SOME EXPERIMENTS ON THE PARANORMAL COGNITION OF DRAWINGS

With special reference to
personality and attitudinal variables

MICHAEL A. THALBOURNE

Doctor of Philosophy Degree

University of Edinburgh

1981

D E C L A R A T I O N

I declare that this thesis is my own work

In memory of my grandfathers,

Christopher Augustine Thalbourne and Laurence Wiesner

A C K N O W L E D G E M E N T S

It is a very pleasant duty to acknowledge my debt of gratitude to all those who have helped me in my parapsychological research. First and foremost, I wish to express my sincere thanks to Dr. John Beloff for his excellent supervision, his unfailing generosity, and his constant encouragement. I would also like to thank my former supervisor, Dr. Peter Delin, of the University of Adelaide; much of this thesis contains the fruits of ideas which his enthusiasm implanted in my mind.

My colleagues in the Parapsychology Laboratory have also generously provided me with advice and assistance whenever I needed it, and I would like to thank them all by name: Dr. Richard Broughton, Hilary Henegan, Kathy Wilson, Yumin Tchen and Deborah Delanoy. Numerous other persons deserve mention for their aid in various capacities, such as Bob Willson, Dr. Peter Caryl, Margaret Menadue, Dr. Erlendur Haraldsson, Dr. Ian Stevenson, and Betty Markwick. A special mention should be made of Tony Burkitt and Dale Stout, each of whom kindly helped with the onerous task of proof-reading the manuscript, and of 'Smokey' who typed it.

I wish to record here my appreciation to the University of Adelaide and the Society for Psychical Research for awarding me, respectively, a George Murray Scholarship and a Research Grant to support me financially during my time in Edinburgh; the British Council, too, are to be thanked for the Commonwealth University Interchange Scheme Grant that enabled me to travel from Australia to the United Kingdom.

But none of this work would have been possible had it not been for the generous cooperation of the several hundreds of people who participated in my experiments either as subjects or as judges. I am most grateful to them all.

L I S T O F C O N T E N T S

L I S T O F T A B L E S

Page

LIST OF FIGURES

A B S T R A C T

The four studies reported in this dissertation were all investigations of the conditions influencing the occurrence of 'general extrasensory perception' (GESP) between pairs of persons. All studies employed the drawing-reproduction technique, in which a person designated as 'agent' concentrates upon a line-drawing depicting some randomly-chosen object, while simultaneously a sensorially-isolated 'percipient' attempts to guess the content of this 'target-drawing' and is required to respond by making a drawing of his or her mental impressions. A single test consisted of ten such picture-guessing trials, three minutes each trial, with no concurrent feedback to the percipient.

The drawing-technique is almost a century old, and accordingly a review of the more important milestones in the history of its use is given. This thesis also makes a number of contributions both to the methods for objectively quantifying degree of target-response resemblance, and to the techniques for determining statistical significance.

The four experiments collectively enable the drawing of four conclusions: (1) contrary to popular belief, the existence of a close emotional relationship between agent and percipient is neither necessary nor sufficient for the occurrence of GESP; (2) group-performance over the ten-trial ESP-test tends to fluctuate rather unpredictably, although there was some evidence that it conforms to a U-shaped or a linear trend; (3) the scores on an 11-item 'Sheep-Goat Scale' (designed to measure belief

in, and experience of, cognitive psi) gave some significantly positive correlations with GESP-scoring, provided that agent and percipient were not in a mutually close relationship with each other; the claim to have experienced at least one instance of telepathy was the most significant predictor of GESP, thus retrospectively providing some validation for the truth of such claims; and (4) the Sheep-Goat Scale-scores of the agents were more strongly related to GESP-scoring than were those of the percipients, prompting an analogy with goal-oriented psychokinesis; it is suggested that this phenomenon can be subsumed under a more general 'goal-accomplishment' principle, called by the author "psychopraxia".

"Oh, terrestial muse, inspire me, please,
to write clearly, without jargonese,
of technical matters. Show me the way
to communicate what I have to say."

W. Morris Dean

C H A P T E R 1

Introduction and Historical Background

CHAPTER 1

" A room hung with pictures, is a room hung with thoughts"

Sir Joshua Reynolds (1723-1792)

I. The term 'paranormal cognition' was almost certainly first used by the English researcher Whately Carington (1945). It can be considered synonymous with J.B. Rhine's more widely used term 'extra-sensory perception' or 'ESP', which may be defined as the "acquisition of information about an event, object or influence (mental or physical; past, present or future) otherwise than through any known sensory channel"*; the expression thus embraces such phenomena as telepathy, clairvoyance and precognition. This dissertation reports the results of four experimental investigations of paranormal cognition -- four attempts to elicit the phenomenon under laboratory conditions and to study its relationship to a number of psychological variables.

Having thus delineated the subject-matter of this thesis as para-psychological, it would be traditional next to set forth some sort of justification for supposing that paranormal phenomena do indeed occur: as recently as Blackmore (1980), students have felt obliged to offer a veritable history of psychical research, starting with the apparent cases of psi which occur spontaneously and unbidden in the course of everyday living, through the formation of the Society for Psychical Research nearly a century ago, up to the technological sophistication of present-day parapsychology. But as Kuhn (1970) points out, it is the mark of an undeveloped science that each researcher

---

* See A Glossary of Technical Terms Used in Parapsychology, Society for Psychical Research, London, in press. This glossary was compiled by the author and a copy of it can be found at the end of this dissertation, as a supplement. It may be consulted if the reader is in doubt as to the meaning of any technical parapsychological terms used in this thesis.

should feel "forced to build his field anew from its foundations" (p.13).

However, during the time that the present writer has been a doctoral student, there have occurred two publishing events which might be thought to intimate the possible demise of this 'pre-paradigmatic' stage. The first of these events was the appearance, in 1977, of the monumental Handbook of Parapsychology (Wolman, 1977), while the second event was the publication of two volumes inaugurating the ongoing series Advances in Parapsychological Research (Krippner, 1977, 1978). Within the covers of these three encyclopaedic works are contained the most up-to-date summaries and evaluations of what is known in parapsychology. While the much hoped for 'repeatable experiment' (looked forward to as the Messiah of parapsychology) is still not within our grasp, yet one cannot but help feel, after studying the above-mentioned works, that nevertheless some progress is being made, be it ever so slowly, towards more accurate prediction of psi.

In particular, the present writer has been much influenced by the argument propounded most forcefully by John Palmer (1977, p.176; 1980), and which may be called the 'probabilistic' or 'pattern-analytic' approach. Palmer argues that given the typically low reliability of ESP-scores, we should not expect one-hundred percent repeatability for an experimental paradigm purporting to demonstrate a relationship between ESP and some other variable, call it 'X'. Rather, we should expect that, if there is really no relationship between ESP and X, any representative sample of studies attempting to establish such a relation will yield (i) an approximately equal number of experiments showing positive and negative relationships between ESP and X, regardless of the significance level; and (ii) an approximately equal number

of statistically significant relationships between ESP and X in each direction (i.e. positive and negative). Says Palmer (1977, p.176): "substantial departures from either of these patterns would suggest that there is a genuine and generalizable relationship between ESP and variable X, even though only a small proportion of the sample relationships were statistically significant, a likely possibility considering the low reliability of ESP-scores". Thus, the pattern-analytic approach involves a comparison of experiments with respect to the direction of the findings rather than to their significance per se. Adopting this perspective, Palmer finds a quite reasonable level of consistency in the experimental literature, whereas previous commentators, (e.g. Rao, 1974), deploring the paucity of significant results, and ignoring trends apparent from non-significant studies, have thrown up their hands in despair.

The present writer is therefore proposing to dispense with a lengthy prologue in the apologetic tradition, and to proceed in the setting forth of this research as if there were a certain minimal consensus as to the existence of at least some genuine paranormal phenomena, even if not well understood or readily reproducible. But this should by no means be taken to imply that the conclusions of parapsychologists have been accepted by (or are even acceptable to) the majority of scientists in more orthodox fields. Far from it! There was, very recently, a proposal made to the American Association for the Advancement of Science, that the Parapsychological Association be "thrown out of the workshop of science" (Wheeler, 1979). Again, Gibson (1979) argued passionately that university researchers be prevented from engaging in the "royal nonesuch" of parapsychology. However, at least one survey of scientific opinion, conducted by the journal

New Scientist (Evans, 1973), has suggested that such dissenting voices are increasingly in the minority, and that a good deal more tolerance is felt towards the "stepchild of the behavioral sciences" (Ehrenwald, 1980, p.116).

The remainder of this introductory chapter will therefore be devoted to a discussion of the particular methodology adopted by the experimenter in his study of paranormal cognition, together with a review of the more important experiments that have likewise made use of this technique.

II. The so-called 'drawing-reproduction', or 'picture-guessing' technique, is one of the classical and time-honoured methods of testing ESP.* In an experiment of this kind, the percipient is required to guess, and draw his impressions of, the contents of a target-drawing, which depicts some configuration of lines or some object selected by a random process. The target-drawing may be concealed, as in a test of clairvoyance; or it may be viewed by one or more agents, as in a test of telepathy or general extrasensory perception.

The drawing-reproduction technique is an example of the "free-response" methodology. Most experiments in ESP research, with the possible exception of physiological studies, can be classified as belonging to one of two broad, methodological categories, of which the free-response method is one. The other general approach that the experimenter may adopt is the so-called "forced-choice" method: in this, the percipient knows in advance that the object or target which he is required to "guess" will be one of a limited range, such as a number or a card from a particular sort of pack. In the traditional

---

* Much of the material in this section has been published as Thalbourne (1979b)

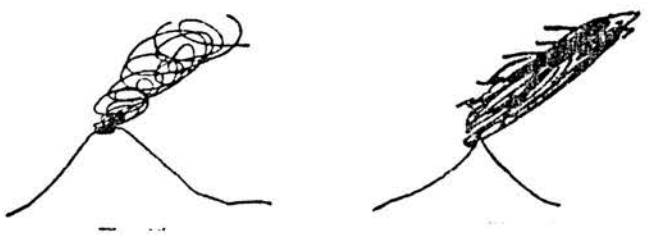
card-guessing experiment, for example, the subject knows that there are just five different targets, namely circle, star, waves, cross and square; and in his response he is, as it were, "forced" to choose between one of these. In the free-response method, on the other hand, the range of possible responses that the subject can make is relatively unlimited. As Carington (1940a, p.38) puts it, the problem for the percipient is not so much "Which?" but rather "What?": he may be asked to record his impressions of what is in the agent's mind, or what is the subject-matter of a book or picture; and this leaves a field of almost boundless extent open to him.

What are the advantages to be gained from using a free-response method rather than a forced-choice technique? To begin with, we may, along with Carington (1940a, p.39), suggest that if paranormal cognition occurs at all, then it may well be to some considerable extent influenced by the degree of interest, emotion, or suchlike associated with the material used. The free-response method is generally found, not only by the agent and percipient but also by the experimenter, to be a very much more interesting sort of task. Again, the method requires comparatively few trials for statistical significance, and so subjects may not become as fatigued or as bored as they do in a long series of repetitive forced-choice guessing. Thirdly, a picture-guessing situation or the like is a closer approximation to the sort of thing that seems to occur in spontaneous ESP: namely, the acquisition of ideas and images -- of whole, original patterns of thought -- and so on, rather than the relatively colorless symbols used in card-guessing.

Another advantage --- again suggested by Whately Carington (1940a, p.39) -- is that with many people, if not with all, an ESP impression which might have difficulty in reaching consciousness in its original and undistorted form might readily appear in a disguised or symbolic shape. When using drawings as the response, some pictorial element of the target-drawing might be recognizable, or the reproduction might represent something more or less closely associated with the original; thus, given a method of assessment which could take account of such modifications without abandoning objectivity, it might be possible to detect genuine paranormal cognitions which would otherwise escape notice. The free-response method can indeed allow for varying degrees of success in any given response: instead of the information present in the response being simply 'right' or 'wrong' -- a hit or a miss -- it is possible to locate it along some sort of scale, and to say that it is more or less informative to a specified degree.

Finally, free-response material often enables us to formulate hypotheses about the content of a paranormal cognition: that is, what is being acquired, and what omitted, distorted or added. In some drawing experiments, for instance, most notably the Upton Sinclair (1930/1962) series, it has happened that the agent's drawing was practically duplicated in regard to its formal, topographical aspects, but the percipient attached a wrong name to it. An example is shown in Figure 1.1, in which the target was a smoking volcano, and the subject, Sinclair's wife, drew a beetle:

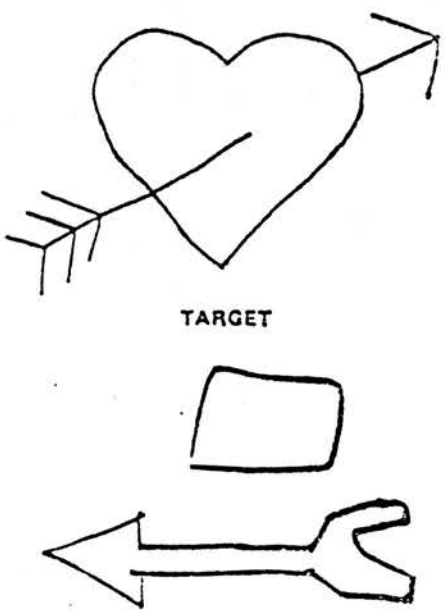
Figure 1.1. Figures 25 and 25a, from Sinclair (1962, p.52)



Could this be misinterpretation of a quasi-perceptual signal, due to a 'noisy' information channel, as when we have difficulty in recognizing objects in twilight? On a very small number of occasions, the word (or homonym) seems to be transferred, and to be misunderstood by the percipient: an example of this from the present author's research might be the case where the target-drawing depicted a bat (the animal), and the percipient drew a squash-raquet, calling it "a bat".

Again, sometimes the idea in the response-drawing is correct, but quite differently expressed from the original in regard to its formal characteristics -- as if the concept were being apprehended rather than a percept. An example is Figure 1.2, taken from the experiment conducted by Targ & Puthoff (1974a) with Uri Geller, in which the two arrows are quite different topographically:

Figure 1.2 Part of Figure 2c, from Targ & Puthoff (1974a, p.605)



Quite often, too, paranormal impressions seem to occur in the form of associated themes, as when in Targ & Puthoff's experiment with

Geller, the target was a drawing of the Devil: Geller drew a serpent, a symbol representing God, stone tablets incised with the Ten Commandments, an apple with a worm in it, and various fruits (which might suggest the Garden of Eden). A consideration of the response-drawings produced in this experiment led Peter Delin (1977) to formulate a two-channel hypothesis of ESP, in which he postulates the existence of two independent channels of paranormal information-acquisition: one mode delivers purely topographic information, while the other yields deep-conceptual information.

For all these reasons, it would seem worthwhile to make use of the free-response method, even though the advantages must be weighed against the fact that the data yielded are much more difficult to quantify than are those obtained from forced-choice methods.

But it remains to be said why the drawing-reproduction technique was chosen here as the particular free-response method. For certainly, the theoretical hypotheses tested in the course of the research could probably have been as readily investigated using even a forced-choice methodology. The simple answer is that the selection was largely a result of personal interest on the part of the author and his former academic mentor. However, if additional justification needs to be given, it might be suggested that one advantage of the drawing-reproduction method that lends it superiority over other free-response techniques is the fact that it requires the percipient to make a pictorial record of their impressions: this record may bear a purely topographical resemblance to the original which, owing to the lack of conceptual similarity, might escape detection if there were only a verbal response to go on. A beetle and a volcano seem to have little in common semantically, but the Sinclair experiments demonstrate that the <u>drawings</u> of each object may be remarkably similar

(see Figure 1.1. p. 6). It is notable in this connection that the researchers Targ & Puthoff, in their most recent studies of 'remote viewing' (e.g. 1978; Puthoff et al., 1980; Targ et al., 1980) have taken to encouraging their percipients to make drawings in addition to verbal reports of their imagery, lest valuable 'formal' information slip through the net.

III. As will become clearer in the prologue to Chapter 3, the experiments to be reported in this dissertation were not tests of hypotheses that were derived from any formal theoretical superstructure: they were, rather, the embodiment of certain ideas that may be traced either to popular beliefs about the conditions favourable to ESP, or (especially in the later stages) to empirical suggestions - no matter how counter-intuitive - that could progressively be gleaned from the author's own data. To the extent that these hypotheses happen to be conceptually related to previous parapsychological research, note will be made of this fact. But it is important to emphasize that the author's own inspirations and deliberations (aided by discussions with colleagues) were the principal forces guiding the selection of research objectives.

But before proceeding forward to narrate the dissertation experiments, it is appropriate to look back at the longer tradition of which they form a part. And what a long tradition that is! At the author's most recent count, there are no less than 71 published reports of experiments which in whole or in part have employed the drawing-reproduction technique -- 17 of these studies having been conducted before the turn of the century! Obviously, it is beyond the scope of this thesis to recount the details of all these studies.* In this brief introduction, then, only a selection of these experiments

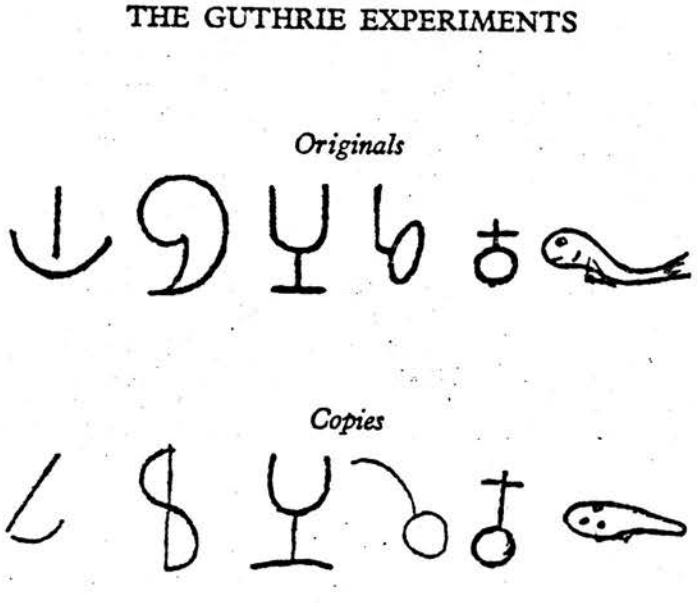
---

* They are, however, listed in chronological order in Appendix 21.

will be given: particular attention will be paid to those studies which constitute the 'highlights' of drawing-reproduction research, either as affording excellent evidence for paranormal cognition, or as having had a major theoretical or historical impact upon parapsychology as a whole.

One of the earliest, and apparently most successful, picture-guessing experiments, was the series conducted by Malcolm Guthrie (1884), in which four different agents and two percipients took part. In some cases, agent and percipient were not previously known to each other -- a fact which reduces the plausibility of the suggestion that the two could have been in collusion with each other. Specimens of their results are shown below, in Figure 1.3.

Figure 1.3. Figures Nos. 1 through 6, from Guthrie (1884, pp.33-35)



A good many of the response-drawings produced in the course of these experiments display a degree of similarity to their targets quite comparable to that of the third pair. Thus, even without the aid of statistical analysis, it was reasonable for the experimenters

to conclude that if the precautions against sensory leakage were adequate, the tests provided good evidence for paranormal cognition.

A large number of experiments similar to those of Guthrie were reported in the early days of psychical research (see Appendix 21, entries 1 to 17), several of these being conducted by such notables as the physicist Sir Oliver Lodge, and the physiologist Charles Richet. However, by modern standards of parapsychological research, the majority of these experiments are generally considered to suffer various defects of design (Thouless, 1972, pp.41,46; West, 1962, p.101). One of the most important defects is the fact that very often agent and percipient were in the same room, thereby rendering it difficult to rule out completely the possibility of conscious or unconscious sensory communication. West, for example (pp. 70-71), in a discussion of spiritualistic seances and stage telepathy, has documented the possibilities of transmission of tiny sensory cues from 'audience' to 'performer', without either party necessarily being aware of what is happening. This is not to say that ESP definitely did not take place in these early researches, or to suggest that the results are entirely worthless. Rather, one feels uneasy at the lack of stringency in certain of the experimental conditions: it would have been much more satisfactory and impressive had the subjects been more sensorily isolated from each other than was thought necessary in those days.

The early experiments are considered to contribute nothing of any weight to the evidence for the reality of ESP as this evidence stands at the present time (Thouless, 1963, p.26). They are of some historical interest inasmuch as they encouraged the early researchers to believe that telepathy was a fact. On the other hand, however,

they suggested that striking results can easily be obtained by the use of drawing-tests, and this suggestion is, on the whole, contradicted by the experience of the last 98 years, at least when using unselected subjects. Furthermore, they enabled the experimenters to avoid the very real problems of objectively quantifying degree of success.

No introduction to picture-guessing would be complete without mention of the outstanding Polish clairvoyant Stefan Ossowiecki (b.1887-d.1944). A chemical engineer by trade, he has been described as a "sealed letter reader",* but by the account of Borzymowski (1965), his paranormal gifts were much more far-ranging and versatile than this. Nevertheless, two of the most famous tests with him did make use of enclosed drawings. The first of these was conducted during the second Congress on Psychic Research held in Warsaw in 1923, and was instigated by Eric Dingwall (1924) -- who at that time was research officer of the London Society for Psychical Research (SPR). Dingwall had written at the top of a sheet of paper the words: "Les vignobles du Rhin, de la Moselle et de la Bourgogne donnent un vin excellent." On the lower half of the sheet he drew a very rough design meant to convey the idea of a bottle. He then placed the folded sheet into an opaque red envelope. This he put into an opaque black envelope and the black one into a brown one, which he sealed. Lastly, he pricked holes through them in four places. This elaborately packaged and sealed target was given to Ossowiecki in bright light in the presence of Dr. von Schrenck-Notzing, Gustave Geley and René Sudre. (Dingwall, very properly, was not present at the critical session). The clairvoyant successfully reported the

---

* Eric Dingwall, in Borzymowski (1965, p.259)

existence and color of the innermost envelope, plus the fact that there was a sketch of a bottle and a French inscription (which he could not manage to read). Dingwall subsequently opened the packet in view of the other investigators and pointed out the precautions he had taken against its being tampered with, adding later (in writing) that these precautions ruled out any possibility of a normal explanation.*

An even more impressive test was carried out a decade later, by Theodore Besterman (1933). Besterman -- research officer of the SPR -- sent from England a package consisting of an outer Manila envelope sealed in a special way with surgical tape, inside which were two additional envelopes, one red, one black, each with "private and invisible marks"; tightly fitted within the innermost envelope was a piece of folded paper on which Besterman had made an outline drawing of an ink bottle, with the word SWAN (underlined in blue) on one side and INK (underlined in red) on the other. Lord Charles Hope, with this in his pocket, tested Ossowiecki in Warsaw. Over a period of half an hour, the sensitive made three drawings, each successively more accurate than the last; the final drawing reproduced the target with almost perfect fidelity (apart from underlining the word SWAN in red instead of blue.) As Hitching (1980) comments, "unless one supposes wholesale collusion, this was, as Besterman reported, a 'brilliant performance'". Tragically, Ossowiecki died at the hands of the Nazis, during the Warsaw Uprising of 1944.

Meanwhile, on the other side of the Atlantic, a series of tests was reported by the American novelist Upton Sinclair (1930/1962),

---

* Beloff (1973) reports that Dingwall eventually came to retract this conclusion, owing to his increasing doubts about the competence and reliability of his collaborator Schrenck-Notzing (who had been in possession of the target-packet for a few hours prior to the test.) Such is the power of the "law of cognitive dissonance"!: "In the fullness of time, every miracle ... will lose its luster." (Beloff, 1973_a, p.195)

whose wife, Mary Craig (b.1883-d.1961), by all appearances, seems to have been an extraordinarily gifted ESP subject. Some of these experiments were carried out at a distance of some 40 miles -- when Mrs. Sinclair was at her home in Long Beach, and her brother-in-law Robert Irwin was in Pasadena, selecting the target-drawings. The resemblances between target and response were frequently unmistakable; indeed, Dr. W.F. Prince (1932/1962) went to the trouble of conducting 'control' tests, by which he demonstrated that the drawings made by supposedly non-psychic subjects bore considerably less resemblance to the targets than did Mrs. Sinclair's. An objection that has been raised to these experiments is that the sender did not choose the targets by any random procedure, simply putting down whatever came into his head, so there is at least a "theoretical possibility that common associations" in the minds of the agent and Mrs. Sinclair "caused them to make similar drawings without the intervention of telepathy" (West, 1962, p.101) (This methodological defect is also shared by the nineteenth century experiments.) This suggestion is not entirely lacking in plausibility in cases where the response duplicates the concept depicted in the target, as when the target drawing was of a fork and Mrs. Sinclair reported that all she could see was "a table fork" (Sinclair, pp.6-7). But as Prince (1962) points out (pp.217-220), (albeit in relation to another rival counter-hypothesis, namely that of "unconscious whispering"), this explanation seems rather otiose when it comes to the large number of purely topographic resemblances (such as the volcano/beetle case), where there is practically no ideational connection with the target.

Be this as it may, the tests impressed no less a person than Albert Einstein and persuaded him to take a somewhat less skeptical

attitude towards ESP than the one he had adopted concerning statistically-evaluated card-guessing experiments (see Ehrenwald, 1978). The Sinclair work is, furthermore, of considerable historical importance in that it also attracted the interest of the psychologist William McDougall: partly as a result of personal testing of Mrs. Sinclair, he decided to set up the Parapsychology Laboratory at Duke University (North Carolina), with young J.B. Rhine as his assistant, and later director.

Standing halfway between the spontaneous cases of real-life psi, and the controlled sterility of Rhine's card-guessing experiments, is the monumental work of the French parapsychologist René Warcollier (1921,1938,1948/1963). Warcollier saw his task as being the qualitative definition of the nature of the material which may at times appear in telepathic communication. A distinguishing feature of his approach is that he considered as evidence of telepathy target-response correspondences which were not literal or exact; indeed, he believed that telepathy was a 'primary process', and that as such, transformations and alterations in imagery should be quite typical occurrences. It was these changes and distortions that he wanted to study and analyze. His work is to be remembered for the many "clinically rich" observations (Cook, 1980, p.97) he offered about the psi-process, drawing parallels as he did between ESP and ordinary sensory perception; indeed, he came to the conclusion that the laws of normal and abnormal perception seem to apply to telepathy. The renowned psychologist Gardner Murphy even went so far as to suggest that the value of Warcollier's experiments lay in the fact that they threw light on important laws of thought and perception in general.

But it is to Whately Carington (1940a,b; 1941; 1944a,b; Carington & Heywood, 1944) in England, and Charles Stuart (1942, 1944, 1945a,b; 1946; 1947; and colleagues, 1947) in the U.S.A. that the credit must go for bringing drawing-reproduction tests all the way into the experimental laboratory, as a result of their (independently developed) methods of statistical evaluation. We shall have more to say in the next chapter concerning their respective innovations in the assessment of free-response material. Suffice it to say for the moment that their techniques may each be described as 'matching' methods. West (1962) succinctly describes the essence of these methods, as follows:

> "Suppose there is a series of ten target drawings, used on ten different occasions. On each occasion a group of subjects has tried to reproduce the corresponding target drawing. An independent judge, ignorant of the order of the target drawings, is asked to compare or 'match' all the subjects' drawings with all ten targets, noting as many resemblances as he can find. The judge's matchings are then analysed to find whether the subjects give resemblances to the target at which they are aiming more often than to the other targets. If there is a bigger proportion of resemblances between subjects' drawings and the target of the same occasion than between the subjects' drawings and the other nine targets, this is evidence of ESP." (pp.101-102)

The basic experimental procedure that Whately Carington (1940a) used was to hang up one of a series of ten drawings inside his (locked) study in Cambridge, at 7 pm on ten successive evenings. Each target-drawing depicted a single object which had been chosen at random from a dictionary. The drawing would remain displayed until 9.30 am the next morning. Between these hours, his percipients -- all living at a distance, and sometimes even in another country -- had to draw what they supposed the target-object to be. Such a series of ten evenings constituted a single experiment. After a period

of time, the ten-trial procedure was repeated, using a new set of targets and a different group of subjects. When a number of experiments had been completed, the drawings produced in the course of the entire series were randomized and sent to an outside judge, who was to score each response-drawing against every target-drawing used. The judge was given no normal means of discovering which drawings had been made in response to which targets; nevertheless, statistical evaluation of his matching-scores indicated that the subjects obtained more 'hits' on the target-drawings in the experiment in which they took part, than on the target-pictures used in the experiments in which they had not taken part. The evidence for paranormal cognition was in fact highly significant.

One of Caringtons most ardent desires was to find a 'repeatable experiment' -- that is, one which any competent investigator could duplicate using a similar group of subjects and obtain comparably significant results. Partly for this reason, he applied his tests not to highly gifted subjects, but to ordinary persons who made no claim to psychic talents. Because he had not selected his subjects in any special way, and because nonetheless he had obtained positive results in several series of tests, he thought that other experimenters would be able to do the same. "I have no doubt at all," he wrote (1944a, p.106), "that the drawings-technique as a whole, using any statistically valid method of assessment, is truly repeatable, in the sense that anyone who cares to do what I have done will obtain substantially the same results; though he may not, of course, if he elects to do something different."

Did subsequent research vindicate Caringtons conclusion? The general consensus seems to be that it did not. For, while the overall

results obtained by Schmeidler & Allison (1948) were remarkably similar to Carington's, those of Taves, Murphy & Dale (1945) were only suggestive and not confirmatory; indeed, six experiments carried out by West (1947) showed no reliable signs of ESP. However, in all due fairness, Carington did explicitly warn that he could not vouch for the success of experimental designs which departed from his own rather strict specifications regarding sample-size (namely, about 100 naive subjects) and number of targets (namely, 10); the studies carried out by West, in particular, departed radically from these specifications. In fact, as we shall describe more fully below, Marsh (1958, 1962; Fisk, 1960) conducted a very large study in closer conformity to the Carington mould, and obtained highly significant evidence for ESP.

Nevertheless, Carington's design was extremely laborious in regard both to acquisition and to evaluation of the data, and it is probably this feature more than anything else that has deterred experimenters from attempting to repeat his work. In the long run, his experiments were important mainly in drawing attention to a curious characteristic of the ESP process known as "temporal displacement effect" -- 'displacement' because it refers to the phenomenon of obtaining hits on targets other than the contemporaneous, or intended, one; 'temporal' because these 'other' targets were either ones already used in the experiment, or ones that were going to be used later in the experiment but which had not yet been selected. A subject might, for example, produce a drawing on Wednesday that showed a striking resemblance not to Wednesday's target-drawing, but to the one that had been used on Tuesday; alternatively, yet

as frequently, it might display an equally notable similarity to the target-drawing which (so it would transpire) would be exposed on Thursday, even though Thursday's target had not, at that time, been drawn and its subject-matter not even selected. The displacement effect thus appeared to involve precognition of a future target, and delayed telepathy (or perhaps clairvoyance or retrocognition) of a past target.

Carington is often considered to be the 'discoverer' of this effect, as if prior to his research it had been totally unsuspected. However, earlier workers had anecdotally reported very similar phenomena: Guthrie (1885, pp.426-7,442) refers to apparent displacement onto previous targets as 'reminiscence', while Sinclair (1930/1962, chapter 14,) used the term 'anticipations' to describe a response matching a target which was still to be used. Nevertheless, Carington is to be credited as being the first to obtain objective statistical evidence that hits of this peculiar kind were occurring to an extent significantly greater than wasattributable to chance. A not dissimilar effect was found much later by Marsh (1958).

On the other side of the Atlantic, and independently of Carington, that other great doyen of the drawing-reproduction technique Charles Stuart (1945b, 1947) also found experimental evidence for displacement. His findings were, however, somewhat different from Carington's: his subjects tended to obtain significant scores exclusively on either the target one behind or on the target one ahead (but not on both); in his 1947 study, this tendency was found to be wholly attributable to subjects who reported feeling 'rushed' or 'inhibited'. John Palmer (1978, pp.198-201) gives a good summary of the evidence

for all these sorts of displacement effect. Though it must be admitted that we have not progressed very far at all in our understanding of why such effects should occur, there seems little doubt that they are genuine phenomena, manifesting not only in free-response settings but also in forced-choice experiments such as card-guessing. It is perhaps not overstating the case to say that had it not been for drawing-reproduction research, displacement effects might have gone unnoticed for much longer than they did.

Nevertheless, Stuart's research is best remembered not for his empirical studies concerning the psi-process so much as for his invention of the 'Preferential Matching Method' (or 'preferential ranking technique', as it is often called today). We shall have much more to say regarding this evaluative-tool in the following chapter on methodology, but it is worth recording here that Stuart's method of assigning a numerical score to the subjective degree of target/response resemblance is, with minor refinements, still in widespread use amongst modern free-response researchers.

With this methodological innovation, Stuart was able to demonstrate the presence of ESP in a number of experiments (e.g. 1942,1945a, 1945b,1946,1947). Even more traces of psi were brought to light from these studies as a result of a famous secondary analysis carried out by Betty Humphrey. The 1940's witnessed the emergence of interest in discovering personality characteristics that would distinguish subjects who 'do well' in ESP tests (we would nowadays say 'who score above chance') from those people who score at or below chance. Humphrey was one of the pioneers of this approach. She informally analyzed data from several subjects who had participated in Stuart's experiments: she compared their ESP-scores with scores derived from

applying to the ESP-drawings four rating-scales developed by the psychologist Paula Elkisch (1945) as measures of maladjustment. Humphrey found that the 'expansive-compressive' dimension seemed to be a good predictor of ESP-scoring, whereupon she embarked upon a formal re-analysis of all available data. Expansive drawings are those which are bold and uninhibited making full use of the space provided, and representing their topic clearly regardless of the artistic ability of the drawer. In contrast, faint, timid drawings, or tiny drawings squashed into a corner of the page, or very conventional drawings indicating that the subject has not given free rein to their imagination, are all classified as 'compressive'. It has sometimes been said that the expansive/compressive dichotomy corresponds roughly to the well-known division of persons as extraverts or introverts. Palmer (1977), however, tends to regard graphic expansiveness as a transient state rather than an enduring characteristic: ".... results with this measure in the context of ESP testing suggest that it lacks sufficient stability over time to be clearly interpretable as measuring a personality trait, and it seems more reasonable to interpret it as a reflection of mood and reaction to the test situation ...." (p188).

Be that as it may, Humphrey took the four drawings made by each subject in certain of Stuart's experiments, and rated each picture separately as expansive (this being given a score of 1), compressive (score of 0), or indeterminate (0.5), these ratings being made before she knew what ESP-scores the drawings had been awarded. A total score was obtained for each subject, and where this score exceeded 2.0, the subject was himself described as 'expansive', all others being 'compressive'.

Humphrey's first report (1946a) concerned her re-analysis of the drawing-data which had been produced under clairvoyant conditions (that is, where no-one was looking at the target-picture at the time when the subject attempted to reproduce it). She found that in each of four series conducted under these conditions, the mean ESP-score of the expansive subjects was higher than the mean for compressives: the difference was significant for the four series combined and for the first series separately.

Humphrey next (1946b) turned her attention to the six series that had been conducted under GESP conditions (that is, where an agent had been concentrating on the target-picture contemporaneously with the percipient's attempt to reproduce it, thus permitting either telepathy or clairvoyance (or both) to operate). It was again found that expansives and compressives tended to score oppositely, but in these circumstances it was the compressives who scored <u>above</u> chance and the expansives below: statistical significance was again attained for the group-difference when all six series were combined, and for one series individually.

The overall results thus indicated a consistent interaction between graphic expansiveness and type of ESP test (clairvoyance vs. GESP). Humphrey's work did not show that one group (expansives or compressives) actually 'had' more ESP than the other, but that one kind of personality showed a tendency to respond with the correct target-information, the other kind to respond in such a way as to <u>avoid</u> the target, depending on the mode of ESP being tested. Indeed, according to Louisa Rhine (1967), it was partly as a result of such findings as these that the negative-scoring effect began to be recognized as a genuine paranormal phenomenon, and to be called 'psi-missing' (as a counterpart of 'psi-hitting', or positive-scoring).

Humphrey's success sparked off a large number of attempts to replicate and extend the predictive power of expansion-compression. Two series were conducted by Bevan (1947a,b), and a further two series by Stuart, Humphrey, Smith and McMahan (1947): trends in the predicted direction were found in all four series, significantly so in two. On the debit side, non-significant reversals of the trend were found in the clairvoyance experiment of Nash & Richards (1947) and for both clairvoyance and GESP by West (1950). In all, there have been 16 series using drawings as ESP-tests, and in 14 of these the expansive-compressive differences have been in the predicted direction, four of them significantly so; in only two were there reversals of the expected trend, and neither was significant.

Not infrequently, graphic expansiveness has also been used to predict the results of card-guessing tests administered in the same session as the drawings, the latter in some cases not being used as ESP-tests themselves. Out of ten series of this type, there have been eight confirmations of the predicted trend: McMahan (1946), Smith & Humphrey (1946: three series reported), Bevan (1947a), Stuart et al. (1947), Kahn (1952), and Kanthamani & Rao (1973); two of these series (McMahan; Kanthamani & Rao) were in fact significant. Thus, of ten cases, only two were disconfirming (namely, Stuart et al. (1947) and West (1950)) and neither of these reversals was statistically significant.

If we accept the logic of Palmer's pattern-analytic approach, described earlier in this chapter, we should expect that if there were no genuine relationship between ESP and graphic expansiveness, then there would be an approximately equal number of confirming and disconfirming experiments (regardless of whether significance

was obtained in any): this random state of affairs seems clearly not to be the case for this psychological variable, since 22 of 26 published series favour the Humphrey hypothesis. Equally, on Palmer's approach, we would expect that for a variable unrelated to ESP, the number of significantly confirming instances would approximate the number of significantly disconfirming instances, but again, the data fail to conform to this random pattern, inasmuch as there are six significant confirmations of the expansive-compressive predictions, but not one significant disconfirmation. (The high premium placed on the publishing of significant results regardless of whether they contradict previous findings, is such that we can be fairly confident that the literature is not much biassed in this respect.) Thus, we seem to be justified in concluding that there is a genuine correlation between graphic expansiveness and ESP: there is a fairly consistent tendency for expansive subjects to score better on clairvoyance and compressive subjects to score better on GESP tests.

Examination of the publication dates of expansion/compression studies indicates that enthusiasm for using it as a predictor variable was remarkably short-lived, probably undeservedly so in light of the conclusions drawn from pattern-analysis. We might speculate that the reasons for its decline might perhaps be related to the more general decrease in the use of drawings as a testing method seen over the last three decades. Alternatively, its decline might be due to the difficulty of explaining the cause for the interaction between graphic expansiveness and mode of ESP. Or it may be that researchers have deemed the percentage of studies that yield significant results (namely 23%) to be too low for expansion/compression to be considered a 'replicable' effect, or too low to justify the large expenditure of labor involved in drawing experiments. If this

last possibility has any truth to it, then like so many other chapters in experimental psychical research, this one may simply be yet another object lesson in how preoccupation with significance (and the assumption that non-significant studies tell us nothing of relevance to the hypothesis tested) has blinded us to the strong trends evident when we adopt the pattern-analytic approach. A way out of this dilemma is to carry out a number of experimental series, and predict that a significant number of series will exhibit the predicted trend (regardless of whether significance is attained in a single study). For example, the present author has to date carried out a total of six discrete drawing experiments, all under GESP conditions. We may therefore predict that in a majority of those experiments, the ESP-scores of compressive subjects will be higher than those for expansives.*

The 1940's were the veritable Golden Age of the drawing-reproduction technique, in that it formed a part of the parapsychologist's test-battery more frequently then than it has ever before or since. After Humphrey's work, and the replication attempts to which it led, the drawing technique seems rapidly to have waned in popularity. For instance, Appendix 21 lists only two published reports in the 50's, compared with 25 for the preceeding decade. Nevertheless, a very large and important experiment was being conducted during the 50's by Maurice Marsh, at Rhodes University, in South Africa (Marsh, 1958,1962; Fisk, 1960).

Marsh's work was directly inspired by that of Whately Carington and of Betty Humphrey, as well as that of John Hettinger (1941)

---

* This analysis has not yet been carried out because,like the Rorschach Ink-blot Test, freehand drawings constitute a projective test of personality, and special expertise (not yet available to the author) is required for making attributions of expansiveness or compressiveness. Nonetheless, this avenue remains open for future research.

and Gertrude Schmeidler (1943,1945,1946). His first aim was to ascertain whether his group of 371 subjects (mainly students at Rhodes University, Grahamstown) would be able to reproduce target-drawings made 470 miles away by an agent in Cape Town, by means of ESP. Secondly he wanted to investigate whether any such ESP scoring ability was related to various personality and attitudinal variables. And third, he wanted to examine the 'Linkage-Problem' -- the question of the nature of the interaction between agent and percipient in an ESP event, and how this interaction can be facilitated. He decided to investigate this problem by objectively testing the efficacy of so-called 'psychometric links' (that is, personal articles which had belonged to the agent, and which are presented to the percipient).

The experiment was divided into five weekly sections of five days each, making 25 experimental days in total, with each of the 371 subjects participating throughout. Each day, the agent randomly selected from 100 possible target-titles two to be illustrated, this process ultimately resulting in the production and use of 50 target-drawings; the remaining 50 titles were later illustrated and used as controls. Over 17,000 response-drawings were returned by the subjects, and they were randomized and scored against the randomized set of 100 target-plus-control drawings by three independent and blind judges. In the strictest category of hits (namely, absolutely correct reproduction of the title of the target), it was found that the judges had awarded 133 hits on the experimental targets, as against 37 on the control targets, this difference being astronomically significant. It was further found that the 133 hits on actual targets were distributed evenly throughout the entire 25-day period of the experiment: they showed no tendency to occur more frequently in

the week a particular target was being used than in the other weeks when it was not. Like Carington's displacement effect, these results suggest that ESP does not necessarily hit the contemporaneous target, but may be pre-cognitive and 'post-cognitive' as well.

When Marsh divided his subjects into a 'high-scoring' and a 'low-scoring' group in terms of the number of hits each had obtained, the high-scorers turned out to be significantly more extraverted (as measured by the Bernreuter B_{31} scale) than the low-scoring group. There were also significant differences favouring subjects who believed in ESP and subjects who were well-adjusted. Finally, the subjects were randomly divided into an experimental and a control group, the former being supplied with a description of the agent's personality and a photograph of him, the latter group being given an equated but incorrect set of such 'linkage material': significant improvements in ESP-scoring were observed for the experimental but not for the control group. A similar effect occurred when the experimental subjects were provided with one-inch squares of cut-up handkerchiefs which the agent had been using for several months, as opposed to indistinguishable squares belonging to someone wholly unconnected with the experiment. These intriguing results lead to the inference that the correctness of the psychometric links given to the experimental group was an important contributor to its effectiveness; had its effectiveness depended merely on its suggestion value, then the control subjects (who were led to believe that the material was correct) would also have shown increases in ESP-scoring.

Since Marsh's work, there has been on average only one drawing-reproduction experiment published every two years. In the present author's opinion, the reason for this decline has been the concomitant explosion in diversity and creativity of alternative experimental

techniques that the 1960's witnessed (see, for example, Beloff, 1974a). While the free-response method itself has never enjoyed greater popularity than at the current time, there has nevertheless been a marked shift towards the use of pictorial targets other than drawings -- photographs and art-prints, for instance (e.g. Ullman & Krippner, 1970) -- in the belief that this will provide an even more lifelike and emotionally arousing target-situation. At the same time, parapsychologists have turned considerable attention towards investigating potentially psi-conducive 'altered states of awareness', such as sleep, hypnosis, meditation, sensory deprivation and the like (see, for instance, Honorton, 1977). In experiments such as these, subjects are required to give verbal reports of their mentation, and these 'mentation reports' are then evaluated using matching-methods entirely analogous to those employed in drawing-experiments. As mentioned in the previous section, however, researchers are beginning to realize the value of having a pictorial record of the subject's response as well as their verbal report; the writer would not be at all surprised if, in the near future, experimenters establish it as a routine requirement that subjects provide both graphic and verbal data as a single 'response-protocol'.

But to return to experiments where drawings were the principal mode of response: approximately 50% of latter-day studies have made use of basically unselected subjects (e.g. Beloff & Mandleberg (1967), Hardy, Harvie & Koestler (1973), Thalbourne, 1979a,1980)), whereas in the other 50% experimenters have employed the drawing technique to investigate the supposed paracognitive abilities of a particular individual. For instance, Pratt (1966) tried unsuccessfully to investigate whether any vestige of ESP remained (if ever there had been some) in S.G. Soal's subject Gloria Stewart. Again, Barry (1971) gives an account of some clairvoyance experiments that he and one

Dr. Dufour carried out with a medium, Madame Marie Maire: in her apartment in Paris she was apparently able to describe or make drawings resembling objects selected as targets by the experimenters in Bordeaux. And most recently, Braud, Davis & Wood (1979) obtained significant results when the English psychic Matthew Manning attempted to draw the scene depicted on each of 30 concealed slides. In fact, before closing this Introductory chapter, we should describe just two more studies which no review of modern drawing-tests should omit, both of them experiments with single-subjects: these are the studies by Musso & Granero (1973) and by Targ & Puthoff (1974a).

Ricardo Musso and Mirta Granero conducted an experiment which is considered by many parapsychologists (e.g. Beloff, 1980, with comments by Palmer and by Child) as being one of the best-designed and evidentially impressive in all of psychical research. These two South American workers reported the results they obtained in 15 sessions (six drawing-trials per session) carried out with a gifted subject by the name of Dr. José Baldomero Muratti -- an Argentinian psychiatrist. Muratti seemed unable to score significantly on card-guessing tests, but he produced highly encouraging results in a preliminary test using drawings, whereupon the researchers set up a formal experiment.

It should be mentioned that the subject personally believed that the form taken by his ESP was telepathy rather than clairvoyance. Unbeknownst to him, therefore, Musso and Granero manipulated the conditions so as to be able to test this hypothesis: one condition was clairvoyance (in which the agent neither knew nor looked at the target), while the other two conditions were telepathy, inasmuch as the agent was aware of the target (in one case the stimulus-drawing having been prepared <u>prior</u> to the test, in the other, chosen and drawn on the spot by the agent as the test was taking place). The

resultant 90 response-drawings were evaluated for likeness to their targets by four blind and independent judges, using a sophisticated system of rank-ordering.

The study yielded highly significant results ($p < 10^{-7}$) for each of the three conditions both separately and combined. Moreover, and contrary to what the subject himself would have predicted, the differences between the three conditions were not significant. At the end of each trial (but prior to any feedback), agent and percipient had also made a note as to the degree to which they felt confident that ESP had taken place in that trial; yet there was no tendency for these confidence ratings to be correlated either with each other or with actual ESP success. Such findings as these led Irvin Child (see Beloff, 1980) to comment:

> " Musso and Granero's experiment, in finding that this outstanding subject (a psychiatrist) was mistaken about the general conditions that favored his psi-hitting and also could not recognize the specific occasions of his hitting and missing, confirms a considerable body of evidence about the usual inaccessibility of psi processes to consciousness even in outstanding subjects. Much in modern cognitive and social psychology would reinforce the point that introspective report has very limited value in guiding research." (p.96)

We come, finally, to the much-publicized and highly controversial study conducted at the Stanford Research Institute (SRI) by Russell Targ and Hal Puthoff (1974a). These two physicists were attempting to validate the ostensible psychic talents of the Israeli performer Uri Geller, and amongst other techniques, they made use of a test with drawings. Thirteen separate drawing trials were carried out over the seven days from August 4th to 10th, 1973; on ten of these trials Geller chose to make a response, while 'passing' on the other three. (Targets and responses are contained in Appendix 8). The trials were carried out under various conditions, but in all of them Geller was shielded visually, acoustically and electrically from personnel and

material at the target locale. Contrary to what some critics would have us believe, Targ & Puthoff were well aware of the strategies and precautions which had to be employed to rule out normal modes of sensory communication between Geller and the targets:

> " These included the necessity of generating target drawings out of sight; the requirement that no one in Geller's presence before or during an experiment should have knowledge of the target (to avoid the possibility of subliminal cueing by body language, subvocalization etc.); the elimination of potential confederates from the target area (to prevent radio reception by means of an implanted receiver); the necessity of maintaining silence in the target area (to foil 'bugging' of the target area, a potentially useful strategy on the part of a subject should he have an implanted receiver);.... the absolute requirement that the target responses be obtained from Geller on paper and be in the experimenter's physical possession before the target was revealed to Geller (to prevent post hoc alteration of the data); etc., etc." (Targ & Puthoff, 1978, pp.174-175).

Some of the responses Geller made were quite remarkable in their accuracy: on the second trial, he reported seeing some "purple circles", and drew a bunch of 24 grapes. The target? A bunch of 24 grapes! (See Appendix 8, figure 1,b). All the response-drawings were evaluated for degree of likeness to their targets by two 'blind' and independent SRI scientists, using the Method of Correct Matchings (see Chapter 2). Each of the judges was able to match each of the ten targets to its own response, which is to say, without error -- an event which would occur by chance at odds of over a million to one.

The experimental design has been criticized by Hanlon (1974) as being inadequate to prevent cheating on Geller's part -- in particular, the possibility that Uri had a radio transmitter concealed in his tooth. As support for this hypothesis,

Hanlon claims that "perhaps the most striking factor which runs through all ten pictures, however, is that Uri seems to be drawing neither the target word nor the target drawing. He appears to base his drawing on the words which would be used to describe the target drawing" (p.180). But as Delin (1977) points out: " ... the most cursory inspection of figures 1 and 2 [Appendix 8] suffices to demonstrate the falsity of Hanlon's statement. Only four of the ten responses could reasonably be described as being what someone might draw in response to a description of the target. ...it is difficult to imagine how Geller could have drawn a horse [Figure 1,e] in response to a cue which surely would have included the word "Camel", or how his informant could possibly have avoided using the words 'cracker', 'devil', 'bridge', 'church', or 'heart' in relation to the relevant targets." This point was also made by Beloff (1974,b): for most of Geller's responses, "the correspondence, although sufficient to enable both the two judges to blind-match all ten response-sets with all ten targets without error, is based on quite small, subtle similarities of formal motif, which in no way suggests verbal cue-ing" (p.357). Targ & Puthoff (1974b) defend themselves by re-iterating the strictness of the experimental protocol.

Not only scientists but professional deceivers -- otherwise known as magicians -- have also vociferously criticized this experiment. Nearly eight years on from the original study, the controversy is still with us. Latest news, according to Teresi (1981) is that Geller is offering a $100,000 challenge to the magician James ("the Amazing") Randi (or indeed to anyone else) who can duplicate the results he produced in

the drawing test at SRI. Geller stipulates that the same controlled conditions as were imposed on him must be used, with the same SRI scientists. Randi is reported to have "gleefully accepted the challenge". The writer awaits the outcome with interest, though with a strong personal bias to suppose that if Randi goes ahead with the challenge he is likely to come off a considerably poorer man.

CHAPTER 2

Methodology: Procedure, and Methods of Assessment

CHAPTER 2

ΑΓΕΩΜΕΤΡΗΤΟΣ ΜΗΔΕΙΣ ΕΙΣΙΤΩ*

("NO ENTRY WITHOUT MATHEMATICS!")

According to the traditions which we have inherited known as 'common-sense', there are two principal ontological dimensions along which all discrete events in the universe may be classified: an event may be categorized according to its location in time (past, present or future, relative to a given reference point); and according to the nature of its substance (physical or mental). Presumably, any event can be a 'target' -- that is to say, the object of an attempt to obtain information about it. If that attempt results in the successful acquisition of such information, and if nevertheless the mode of acquisition is inexplicable according to the canons of orthodox science, then we call this event an instance of 'paranormal cognition', or 'extrasensory perception'.

The events comprising the genus paranormal cognition have traditionally been classified into several species, according to the temporal location and nature of the presumed target-event. If the target is located in the present, then the temporal mode of the paranormal cognition would be described as 'contemporaneous'; if the target is an event which has already occurred and now no longer exists, then the mode would be said to be 'retrocognitive'; while if the target-event is one that does not presently exist but will come into being at some point

---

* From Elias, commentary on Aristotle's Categories, 118.18: the words, inlaid with gold, said to have been carved over the portico of Plato's Academy, Athens.

in the future, then the mode would be classified as 'precognitive'. Again, if the target is presumed to be essentially mental in nature (such as the contents of another person's consciousness), then the form of paranormal cognition would be described as 'telepathy'. If, on the other hand, the target-event is physical in nature, then extrasensory perception of it would be said to be an instance of 'clairvoyance'. Conceivably, telepathy or clairvoyance could occur in any of the three temporal modes, thus resulting in six possible distinct types of paranormal cognition. This sixfold typology of the varieties of ESP is illustrated in Table 2.1 (adapted from Beloff, 1974a, p.2):

Table 2.1 The traditional 3 x 2 taxonomy of the forms of ESP, according to temporal location and nature of the presumed target-event.

Nature of Presumed Target-Event	Location of Presumed Target-Event		
	Past	Present	Future
Mental	Retrocognitive Telepathy	Contemporaneous Telepathy	Precognitive Telepathy
Physical	Retrocognitive Clairvoyance	Contemporaneous Clairvoyance	Precognitive Clairvoyance

Whether any or all of these forms actually exist is another question, but at least it would generally be allowed that they are conceptual possibilities, given our ontological categories.

But though it is easy enough to enumerate these varieties of ESP, it is frequently difficult if not impossible to classify unequivocally a given instance of paranormal cognition as being one form operating to the exclusion of all others. Particularly is this so for cases of ostensible telepathy (of whatever temporal

mode): there is usually always, at some stage or other, an independent, publicly available physical record of the mental event which is nominally the target, and the information contained in such records could conceivably be obtained by clairvoyance; furthermore, even in those cases where there is no external physical target (such as in the 'pure telepathy' experiment of McMahan (1946)), the mental processes of the person 'transmitting' the target information will surely be accompanied by (some philosophers would say "will be identical with") certain neurophysiological processes in that person's brain, and again information could be acquired about these brain-states by clairvoyance. Thus, because of the difficulty of excluding the possible operation of clairvoyance from ESP test-procedures, most modern-day experimentalists disdain the word 'telepathy', preferring instead the more non-committal and colorless term 'general extrasensory perception' (GESP). This term is applied to instances of paranormal cognition where there are at least two possible candidates for the title of target-event -- a mental event and a physical event.

The aforegoing conceptual analysis is all by way of introduction to the statement that in each of the experiments reported in this dissertation, the focus of attention was on evidence for contemporaneous GESP in pairs of human beings -- in particular evidence that one person can, by making a line-drawing, demonstrate that they have (paranormally) acquired information about a target-drawing being viewed simultaneously by another, sensorily-isolated person.*

---

* Some methods of evaluating ESP success with drawings -- most notably the method used by Carington (see below) -- yield an

In parapsychological parlance, the 'recipient' of the paranormal information is termed the 'percipient' (even though it is a very moot point whether perception is an adequate analogue for the processes that actually occur in so-called 'extrasensory perception'; probably 'respondent' is the most appropriate word). The person whose role is to contemplate the target-drawing and to attempt to 'transmit' a mental description of it to the percipient, or in some way to cause the latter to respond 'appropriately' to the target-information, is known as the 'agent'. Whether or not the agent actually does thereby 'help' the percipient improve their chances of acquiring the target-information, is one of the great unanswered questions of parapsychology. The widespread use of agent/percipient testing conditions is certainly ample evidence that experimenters believe this to be the case; but studies that have shown GESP conditions to be clearly and unambiguously superior to clairvoyant ones alone, are few and far between (see Palmer, 1978, pp.96-102). Be this as it may, the GESP procedure was necessary in the present series of experiments inasmuch as a variable frequently manipulated with a view to observing its effect upon ESP, was the social-emotional relationship between agent and percipient.

In dealing with the persons actually participating in

---

overall score which reflects the level of paranormal cognition operating no matter what the temporal mode. (One may then proceed to 'dissect' this score into the components attributable to contemporaneous and to non-contemporaneous ESP.) The method of analysis employed in the present report, however, is such that only one temporal mode at a time may be examined, and in this case the choice was the contemporaneous mode. But certainly the possibility remains open that in a future research project evidence could be sought of paranormal cognition operating in modes other than this one.

these experiments, the technical expressions 'agent', 'percipient' and 'GESP' were not used. Rather, in order to present a more intelligible and 'face-valid' situation, the agent/percipient paradigm was described as a test of telepathy; the agent was referred to as the 'sender' and the percipient as the 'receiver', as if telepathy were a kind of 'mental radio'. In fact, these locutions will be encountered fairly frequently in the text to follow; however, the reader should bear in mind that they are being used simply for the purpose of stylistic variation without presupposing the validity of any 'transmissive' or 'radiational' model of agent-percipient interaction.

As, of course, will be evident from the Introductory chapter, the information which the agent in these experiments was required to try to 'send' to the percipient concerned the content of a target-drawing; and the percipient was required to demonstrate, if they could, the acquisition of this information by producing a free-hand drawing of their impressions. These are the basic constituents of a discrete attempt to display paranormal cognition -- a single 'ESP trial'. How various persons were selected to be tested in this way; what characteristics they possessed; and the means by which the roles of agent and percipient were assigned, will be described in the following chapters as part of the details of each individual experiment. But of course many other questions of detail arise: how were the target-drawings selected? how many were there, and how manufactured? how many trials were conducted per agent/percipient team, and how long each trial? All these questions, and more, are answered in this present chapter, which concerns the mechanics and apparatus

of the methodology: its purpose is to provide the reader with all the relevant details of how the experimenter set about collecting and evaluating the picture-guessing data which form the grist for his hypothesis-testing mill.

SECTION I

General description of the experimental materials and procedure.

Target-Selection. In the course of our historical overview of previous drawing-reproduction experiments, it was pointed out that the results of many of these studies can be brought into question on the grounds that an insufficiently random method was employed for generating the targets used. Often, the agent or experimenter drew "the first thing that came into their head". But as Thouless (1972) reminds us,

> "... there are few things less random than the processes of thought, and one cannot make a reasonable guess as to how likely it is that two people will think of drawing the same thing at the same time, especially if they are exposed to the same environment. There may, for example, be a case of mushroom poisoning reported in the newspapers which makes more probable the drawing of a mushroom by both experimenter and subject". (p.46)

One might also mention the response-biases to which people are prone in a free-drawing situation: just as people asked to "pick a number" will tend to say "seven!", or if requested to think of a color, are likely to come up with the answer "red", so too will percipients often draw such objects as a house, dog, tree or flower (which are perhaps the first things that children in Western civilization ever draw). It is very telling, for example, that for an (admittedly informal) preliminary test with their high-scoring subject Dr. José Muratti, the researchers Musso & Granero (1973, p.14) 'chose' six targets,

four of which depicted the above-mentioned objects!

Similarly, the selection process must eliminate effects due to the ideas of agent/experimenter and percipients running in similar cycles. Pratt (1937) suggests a subtle example of this for cases where the agent prepares the targets just before each trial, usually immediately after the results of the previous trial have been examined and discussed with the percipient: the agent might tend to alternate, in his choice of targets, between figures with predominantly curved and predominantly straight lines, and the precipient might unconsciously anticipate these changes. For all these reasons, then, the method of target-selection needs to be as random and as objective as possible.

Whately Carington, in his earlier studies on the paranormal cognition of drawings, selected his originals by opening a dictionary at random and making a drawing of the first depictable object encountered on that page. Targ & Puthoff (1974a) used a similar method for the drawing-test they conducted with Uri Geller. In the issue of Nature in which Targ & Puthoff's report appeared, the editorial was particularly critical of this method, stigmatizing it as a "naive, vague and unnecessarily controversial approach to randomization" (p.559). The editors did not, however, elaborate upon this criticism, nor did they suggest a better method.

Maurice Marsh (1958, 1962; and as reported by Fisk, 1960), objected to the dictionary method on the grounds that there was no control over the 'popularity' of the targets so chosen. Says Fisk: "... if the choice of an original happened to fall on 'car' large numbers of subjects might be expected to draw

cars from chance alone and this chance scoring would tend to obscure any genuine ESP" (p.222). If the present author understands this argument correctly, the argument is not that failure to control for target-popularity leads to spurious evidence for ESP, nor even that the method of statistical evaluation adopted is somehow rendered less powerful. Rather, given that an experiment has yielded statistically significant ESP results (which is to say, evidence for an extra-chance number of correspondences between targets and responses), it then becomes possible to consider the question of which correspondences (out of the total number that are correct) are due to chance and which to psi: a certain number of them will be due to coincidence pure and simple (this number averaging out to MCE), but an excess over this number will be due to ESP. Perhaps Marsh's objection is that indiscriminate admixture of 'popular' and 'rare' targets will make it harder to separate the wheat (i.e. the ESP hits) from the chaff (the chance hits): if, for instance, the target portrays a very unusual object -- such as a spark plug -- and in fact a spark plug is drawn as the response, then this is far more impressive evidence for ESP than if the target had depicted a house and the percipient drew a house (a very frequently observed response). If all the targets are in fact of 'rare' objects, then any hits that we might obtain on them will thus seem much more likely to be due to psi than to chance.*

---

* Marsh's own method of target-selection was as follows. Before starting the main experiment, his 371 subjects (mainly students at Rhodes University, South Africa) were each asked to list 50 objects which they considered could be drawn easily and unmistakably. Nearly 16,000 such suggestions were submitted

But if this is the correct interpretation of Marsh's objection, then it seems to the present author that the argument is not really relevant provided that the only issue in which we are interested is the question of statistical significance (as opposed to the post hoc business of pinpointing 'genuine' ESP-hits rather than 'pseudo-hits'). In the studies reported in this dissertation, the author confined the inquiry to the question of whether the subjects (as individuals or as groups) showed any sign of ESP in their performance over a number of trials, rather than the question of which particular target/response matches were especially notable as ESP hits. The author could therefore see no cogent reason for not using a dictionary technique, and this was the method opted for, the target-selection procedure being as objective and as random as possible.

An encyclopaedic dictionary was used (Hanks, 1971), in conjunction with tables of random sampling numbers (Kendall & Babington Smith, 1939). First, the tables would be used to pick at random a page in the dictionary; then, a coin would be tossed to decide which of the two columns on the dictionary-page to select (heads being consistently the left-hand column, tails the right); and finally, the tables would again be entered so as to select at random an item within the column. The word so selected was illustrated as a target provided that it was (i) a concrete noun, (ii) not a proper noun unless easily recog-

---

by the subjects, and these were examined in order to find 'singles', that is, titles which had been suggested only once in all those 16,000 times. Nearly a thousand of these suggestions turned out to be 'singles', and an attempt was made to compile from these a list of 100 titles as unlike one another as possible, lest any confusion of shape arise. These 100 were illustrated as targets -- 50 experimental and 50 controls. Thus, by choosing as targets objects which had been thought of only once in 16,000 trials, Marsh hoped to reduce to a minimum the hits due to chance, and thereby prevent the 'dilution' of genuine ESP effects.

nizable (e.g. 'Pegasus'), (iii) not a scientific-technical term, and (iv) provided that it was fairly easy to draw and recognize, such that anyone of average drawing ability could produce a passable rendering within three minutes. If an item failed to meet all these criteria (as was frequently the case), then the coin would be spun again to determine whether to progress up the column (if the coin showed heads) or down it (if tails) in an attempt to locate a more suitable item. If the entire column failed to yield an appropriate word for depiction, then the selection-procedure was re-initiated da capo. This, then, was the general manner in which the experimenter chose a word to be portrayed in a target-drawing.

It was decided, admittedly somewhat arbitrarily, that the number of drawing-trials carried out by each agent-percipient team would be ten. In retrospect, this number was chosen most likely because of its inherent 'naturalness' in a decimal society, and because the experimenter's use of the drawing-technique at all was first inspired by Targ & Puthoff's test with Geller, which resulted in ten trials (albeit stretching over a week). Thus, a minimum of ten different target-drawings was needed.

Initially, it was thought that this number would suffice for testing all the subject-pairs required. Now an obvious pre-requisite in any ESP test using free-response materials such as drawings is that the subjects be given no information whatsoever regarding the particular objects depicted in the target stimuli. Were only ten targets to be used for the entire experiment, then there would arise intractable problems of sensory leakage: for agent-percipient pairs were going to be

tested not simultaneously, in a group, but individually, and consecutively, the whole experiment taking some weeks from start to finish. There would thus have been no guarantee that later pairs coming into the laboratory for testing would not have heard from earlier pairs what the targets were. Randomizing the target-sequence for each new pair might have been a way of dealing with this dilemma, but this would have tended to turn the situation into a forced-choice task, as well as devaluing the evidential quality and instructiveness of any hits. A better way of circumventing the problem is to use a different set of targets for every pair; the only real difficulty with this is the practical one, inasmuch as the process of randomly selecting and depicting targets is very arduous and time-consuming, especially considering that a total of 102 pairs ended up being tested in the course of this series of studies! (Moreover, having no outside help, the lot fell to the experimenter himself to select and illustrate all the targets). Thus, (for three of the four studies reported) it was deemed more economical of time and targets to use one target-set per experimental session, and test as many subject-pairs as was practicable within that session, though in such a manner that pairs tested later in the session would have no opportunity to learn, from pairs tested earlier in the session, about the particular targets being used. (More details concerning this scheduling will be given below, in the Experimental Procedure.) In practice, up to four pairs would be tested in a single session.

Prior to conducting the four studies to be described in this dissertation, the author had already employed the random-

selection procedure to produce a total of 120 targets for use in the experiment reported as Thalbourne (1976). These targets were also used in three of the four dissertation studies (Chapters 3,4 and 6). They had each been drawn by the experimenter with a black (or, in ten cases, a dark-blue) felt-tipped pen, on A4-size duplicating paper held long-side horizontal. A second party had then screened the finished drawings for quality and suitability.* Ten drawings were rejected as slightly inferior and it was decided that these would be used only for pilot sessions or in case of emergency (e.g. running out of unused target-sets at the end of an experiment). (This set is set 12 in Appendix 1.) The remaining 110 targets had then been randomized (using shuffled counters), and from the resultant random sequence divided into 11 sets of ten, with the proviso that no two targets within the same set were 'overly' similar in concept or form. For the Thalbourne (1976) study, the 120 targets had each been given an identification number ranging from 1 to 120 in systematic order (targets 1 to 10 constituting set 1, targets 11 to 20, set 2, and so forth). Since the targets were to be presented to the subjects in arithmetical order, these identity numbers thus revealed the sequence in which they were used. This fact posed no problems until the study reported in Chapter 6, when 'judges' were presented with complete sets of targets. Although the ten targets within each set were randomized before presentation to the judge, the identity numbers would still have provided a clue (albeit an esoteric one) that could have jeopardized the validity of the assessment procedure.

---

* The author wishes to thank his then-supervisor, Dr. Peter Delin, for performing this task.

It therefore eventually became necessary to whitewash over these original identity numbers and replace each of them with a unique, four-digit code-number chosen at random from the afore-mentioned Tables of Random Sampling Numbers. It then became impossible to deduce rationally which position in the sequence of ten that target had occupied.

For various reasons (set forth in Chapter 5) it ultimately became desirable to select an additional 60 targets, and the procedure for doing this was identical to that described above. Thus, by the end of the dissertation experiments, there was a grand total of 180 targets (that is, 18 sets of ten), and they are listed in Appendix 1. Each target-set was stored in its own Manila envelope, and kept under lock and key until required.

Apparatus and Laboratory Design. The blank sheets of paper on which the response-drawings were to be made were A4-size duplicating-paper held long-side horizontal (the same size and orientation as the target-drawings). Each percipient was provided with 11 sheets of such paper -- one for a practice trial (which was not to be included in the analysis), and ten for the experiment proper. Each response-sheet was inscribed (in the lower right-hand corner) with a unique four-digit code-number chosen from the Random Number Tables (analogously to the targets), again with the dual purpose of being uniquely identifiable (to the possessor of the key) yet at the same time such that no-one could rationally deduce which position it occupied in the sequence of ten response-drawings: likewise, this was methodologically necessary for the assessment procedure

(see SECTION II). The code-numbers of both targets and response-sheets were recorded by the experimenter prior to the test, on a record-sheet specially designed for this purpose, and in the order in which they were to be used in the ten-trial test. (See Appendix 4.) (Note that whenever a particular target-set was used, the order of its constituent targets was always as given in Appendix 1).

The percipient was also supplied with two black drawing-pens with which to make her or his response-drawings (two in case one should happen to run dry during the course of the test). There were also two box-files (foolscap size, and approximately $1\frac{1}{2}$" deep): in one box were kept the blank response-sheets in order, in the other, the response-drawings were to be placed upon completion at the end of each trial.

The agent was likewise provided with two such boxes -- one in which the unseen targets were located, in order and face-down, and one in which to put each target after it had been viewed. However, even though the target-drawings were kept face-down, the paper was not sufficiently opaque that they could not be discerned through the back. Therefore, special card-board 'frames' or 'windows' were constructed (from a Manila folder, trimmed to A4 size, sealed on three edges, and with a quarter-inch border on the open side); each target was placed inside one of these windows, which, when face-down, offered no distraction to the agent; since, too, these windows were numbered I to X, this would act as a safeguard to ensure that the target-order would be adhered to correctly.

Figure 2.1 Plan of the Adjacent-Room Laboratory Set-Up.

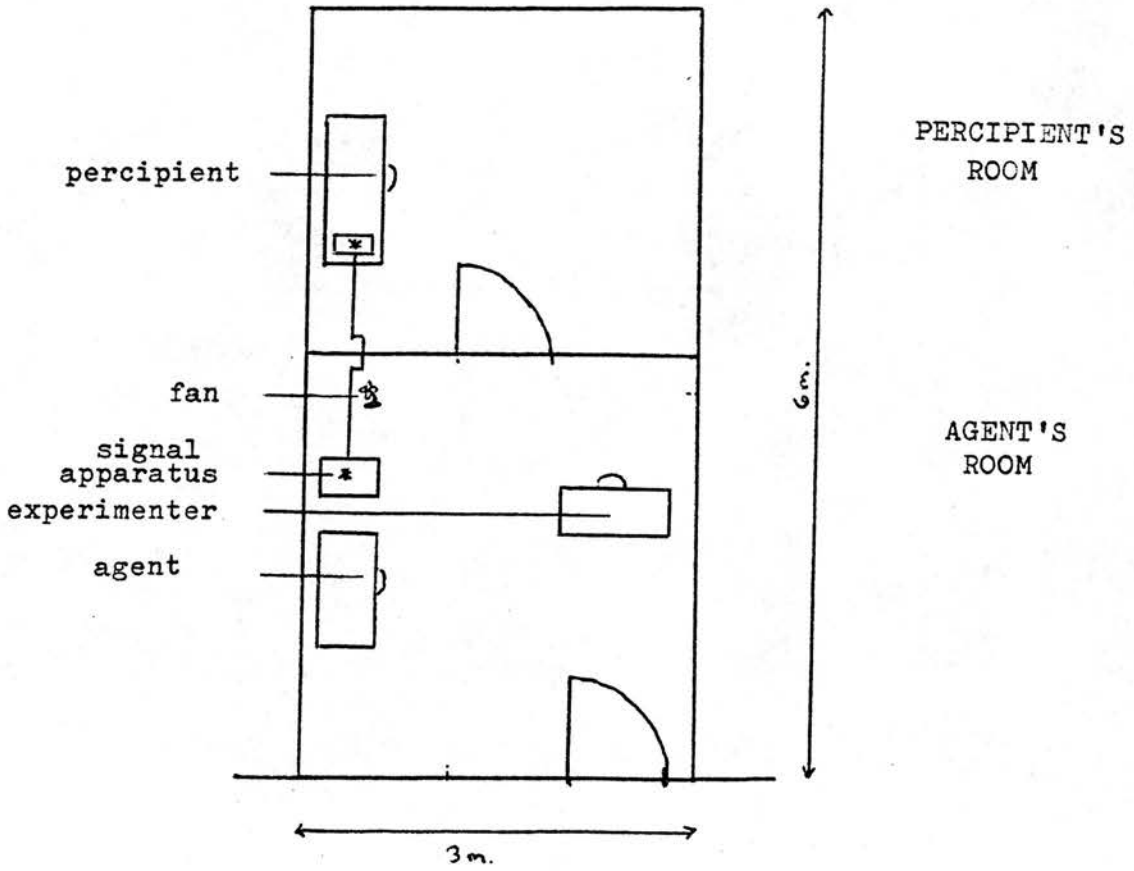


Figure 2.1 shows a plan of the laboratory set-up. Owing to scarcity of research space, adjacent rooms had to be used in the experiments described in Chapters 3, 5 and 6. In the experiment recorded in Chapter 4, two separate rooms were used, distant from each other by about 13 metres. Some readers might feel, especially with regard to the adjacent-room set-up, that sender and receiver were too near each other to rule out adequately the possibility of conscious or unconscious communication (e.g. by inadvertent whispering, loud talking, tapping on the adjoining door or wall, etc.). Though the author would be the

first to admit that the situation was less than ideal, it seems to him extremely unlikely that any such communication ever went on: throughout the ten-trial test, the experimenter would be stationed unobtrusively in the sender's room, and for what it is worth, can honestly state that he never once observed any subject even attempting to communicate with their partner (though of course the experimenter may have been deceived in this matter). Additionally, and to anticipate somewhat the results, most of the evidence for ESP derived from pairs of persons who did not even know each other prior to the test, thus making the existence of a pre-arranged code extremely unlikely; indeed, the couples who knew each other very well, and who might therefore have been suspected of collusion, tended to give no evidence of significant target-response resemblances! Again, this thesis summarizes a total of 1020 discrete attempts to reproduce a target-drawing (ten responses by each of 102 percipients), and on none of these occasions has the response duplicated perfectly the concept portrayed: the resemblances seen tend to be more subtle relations of form or of association (examples of the latter being a radio-response to a television-target, or a tea-cup drawn in response to a wine-glass); this fact seems to counter-indicate the hypothesis that cheating took place in any efficient way. Finally, it should be pointed out that evidence was still found for ESP in the experiment where the two non-adjacent rooms were used, and indeed more evidence than in the study reported in Chapter 5 where the adjacent-room set-up was employed!

But all the honesty in the world could not have prevented subjects from communicating 'unconsciously', or 'subliminally'.

As Sidgwick (1896) reported early in the history of psychical research, skeptical psychologists such as the Danes Lehmann and Hansen (1895) could invoke the hypothesis of unconscious whispering in order to explain positive results in telepathy experiments where sender and receiver were in relatively close proximity (in particular, in the same room). Hence, to be on the safe side, a fan was employed in the agent's room so that, when running, its noise would act as a mask for any inadvertently made sounds.

It had been decided that each trial would be precisely three minutes long, making the entire test extend over one half-hour. It was therefore necessary to have some way of keeping agent and percipient synchronized with each other, so that it would be known which response was meant for which target. Some sort of verbal signal given by the percipient might have been acceptable had timing not been so crucial,. but in any case this was scarcely feasible in the 'distant-room' set-up. Thus, a signal apparatus was constructed to perform the task electro-mechanically.* The machine consisted of two pieces of equipment, one part in the sender's room, the other part in receiver's, and linked by an electrical cable. Continuously and automatically, in both rooms, a reddish/orange light would alternately light up for two minutes and then go off for one minute; whenever the light went on or off, a tone would sound simultaneously -- an additional signal in case the light-change was not noticed by the subject. While the light was on, agent

---

* The author would like to express his gratitude to the technical staff of the Psychology Departments at both Adelaide and Edinburgh Universities for constructing apparatus to perform these functions.

concentrated on target, and meanwhile percipient tried to reproduce that target; when the light went off, the agent laid aside the target-drawing and relaxed, and the receiver finished off the response-drawing. Receiver thus had a total of three minutes in which to attempt each reproduction. At the end of the one-minute 'light-off' condition, the light would come on again and the tone sound to herald the beginning of a new trial.

The only points remaining to be said about the equipment used are that both sender and receiver were provided with an ordinary chair in which to sit, and a table at least one metre long and half a metre wide (to accommodate the two boxes, and the signal apparatus, as well as leaving ample room for the drawings to be viewed or made). The illumination of the rooms was uniform and at a level normal for interior lighting. The walls of the testing rooms were entirely unadorned, lest such decorations suggest ideas to the percipient and thereby render his responding less 'free'.

Apart from the signal apparatus, then, one can see that very little technical gadgetry and only minimal financial outlay is required for drawing-reproduction research -- a distinct advantage in days of stringent economies!

Experimental Routine. Prior to being tested, each subject was assigned a three-digit identity number, so as to preserve their anonymity and to facilitate data-recording. Likewise, each pair was assigned an identity number, decided upon purely by the order in which they were tested: thus, for example, in the Married Couple experiment (Chapter 5), the first pair

to be tested was assigned the number '62', being the 62nd pair ever tested by the experimenter, while the next pair were designated as number 63, and so on.

Upon their arrival at the laboratory, subjects would be welcomed, introduced to one another if not previously acquainted, and have the general set-up of the 'telepathy' test explained to them. Any necessary personal details were then taken, and each subject would typically fill out one or more questionnaires as the experiment required (such as an attitude-to-ESP inventory).

All testing sessions were conducted at the Psychology Department, most frequently during the evening but occasionally on an afternoon at the weekend, so as to ensure that the maximum degree of quietude would be prevailing. Approximately one hour was required to test a single subject-pair. As regards the typical timetable in a given session, it was arranged that up to four pairs would be tested. The first pair of the evening would arrive at the laboratory at (say) 7pm, with the remaining two or three pairs arriving 30-45 minutes later (at a different part of the building), by which time the first pair would be in the middle of the ESP test. These later pairs would have been directed to wait in a comfortable room where refreshments and reading material were available; (to save time, any pre-experimental questionnaires were filled in while waiting). After the first pair had completed all testing, experimenter would debrief them, explaining the purpose of the test in detail, and answer any questions; they would be thanked for their participation, asked not to discuss with anyone what the targets had been, and were personally conducted out of the building so as to prevent them from coming into contact with the subjects

who were still waiting to be tested. The experimenter then fetched the second pair, and repeated the test-procedure anew. Pair three would thus have to wait about an hour, and on the few occasions when there was a fourth pair, the hapless subjects would have to wait for two hours.

This scheduling method was employed in the studies reported in Chapters 3, 5 and 6. But since subjects were unpaid, it was eventually deemed too great an imposition on their generosity to have to ask them to wait like this; also, the effectiveness of the time-tabling depended upon the subjects' punctuality. (Very occasionally, persons would arrive late, and thus could conceivably have colluded with a pair already tested in that session; but even in these cases, the response-drawings they produced never showed any evidence of such ill-gained knowledge: cf. the remarks made above about subject-cheating.) Because of these disadvantages, this cloak-and-dagger routine was eventually replaced by a simpler but almost-as-satisfactory method in which a different target-set was used for each pair tested in a given session (see Chapter 4).

Procedure for the Telepathy Test. Subjects were brought into the telepathy research rooms in their pre-arranged pairs, the fan having been turned on prior to their arrival. The roles of sender and receiver were either assigned by the experimenter (as in Chapter 3) or else allocated according to subject preference. (Once these roles had been assigned, the subject remained in it for the entire test, without switching.) Subjects were then seated at the table in their respective rooms, and then proceeded to familiarize themselves with their written instructions (see Appendices 5 and 6), as well as the target- or drawing-

materials (although the agent was not permitted to look at any of the targets at this stage). The behaviour of the signal-apparatus was explained, after which the single practice trial was conducted (using as target the Star of David). Any problems or misunderstandings could then be rectified.

It should be made explicit at this point that in most of the studies (with the exception of Chapter 4) a slight opportunity for sensory leakage was present, inasmuch as, having no assistant, the experimenter had to prepare the targets by inserting them into their windows and putting them in their required order in the agent's box. He thus obtained some information about what the targets were, and it is conceivable that he could unwittingly have dropped cues to the subjects while interacting with them prior to the test. Of course, these cues would have had to be dropped 'in the right order': even so blatant a cue as the word 'apple' when the target-set in fact contained a drawing of an apple would be to no avail unless there was also some clue given as to which of the ten trials would have the apple as its target. The author can only plead that, as experimenter, he was extremely aware of this possibility, and therefore strove to the utmost to avoid any reference even to what the targets <u>might</u> be. But the one-man laboratory is far from ideal in parapsychological research. It would have been far preferable -- had it been possible -- to employ an assistant to perform such tasks as would ensure that the experimenter would remain 'blind' throughout all his dealings with the subjects. All too true is the statement made by Beloff (1980) to the effect that "in retrospect one can always think of some additional controls one could have introduced or something

one would have done differently" (p.94).

But to return to our description of the telepathy test procedure. After the practice-trial was completed, and any problems ironed out, the test proper was begun: experimenter reset the signal-apparatus, and the pair proceeded to attempt 'transmission' and 'reception' of the ten target-pictures, one target at a time; all the while the receiver was left alone in his room (with the adjoining door(s) closed), and the experimenter would remain quietly in the sender's room so as to keep a watchful eye on the proceedings and make sure that all went smoothly; for most of the duration of the test he was free to read, tabulate data or score questionnaires.

When the three minutes for the tenth and last trial had expired, the two subjects, remaining seated in their respective rooms, were given a post-experimental questionnaire to fill in (Appendix 7): this contained two multiple choice questions (one concerning their belief as to whether or not ESP had occurred, the other, whether they thought there had been any factors that may have inhibited the occurrence of ESP (and if so, what they were)); and five rating-scales enabling subjects to describe their moods and feelings during the test (the bipolar opposites being 'tense-relaxed', 'bored-interested', 'calm-agitated', 'confident-uncertain' and 'enthusiastic-unenthusiastic'). The experimenter then re-entered the percipient's room and went through the response-drawings to ensure that none of the code-numbered response-sheets were blank and that all had been used in the correct order (as recorded on the Telepathy Test Summary Sheet: Appendix 4). Experimenter then took the percipient and their response-drawings to the agent's room: the targets were

compared with their responses in order to satisfy the curiosity of both subjects and of experimenter as to how successful at telepathy the pair appeared to have been. Debriefing having been completed, the subjects were then escorted out of the building.

SECTION II

Methods of Assessment, choice of statistics and tests of significance.

The 'brute' raw data yielded by the above-described procedure consist, of course, of two sets of pictorial matter: on the one hand, the 'stimuli', or target-drawings, each depicting as their subject some arbitrarily-chosen subject; on the other hand, the graphic representations made by a number of persons in their attempts to acquire information concerning those stimuli when outwith the range of sense and rational inference, which is to say, paranormally. The pertinent relation of interest is the 'similarity' between target and response, which in this case was deemed to be the extent to which the 'information' in the response corresponds accurately with that in the target. To the extent that this 'amount' of resemblance consistently deviates from that 'quantity' that would be expected on the hypothesis that the percipient was merely guessing, to that extent we may say that an extra-chance factor is likely to have been responsible for this circumstance. This portion of the chapter is therefore devoted to a review of some of the relevant methods of quantifying degree of target-response resemblance (which is to say, the techniques that have been devised for assigning a numerical score to the psychological property

of 'similarity'); we will then describe some of the statistical procedures employed to ascertain whether or not this quantity deviates significantly from chance.

The Method of Paired Comparison. Early drawing-reproduction studies (i.e. prior to 1929) frequently yielded target-response resemblances of astonishing fidelity; the similarities were often so obvious to unaided inspection that little need was felt for a more quantitative method of assessment. The first researcher to attempt to develop a more objective evaluation was Walter Franklin Prince(1932). His method was quite laborious, and was prefaced on the assumption that paranormal ability is a very rare phenomenon. The technique consisted essentially of repeating the experiment to be evaluated, using the same set of targets but with a group of randomly chosen 'control' subjects (i.e. persons supposedly lacking in psychic ability). A 'blind' judge then compared the responses of control subjects and the response made in the original experiment with the actual target, saying which was the better match (hence 'paired-comparison'). Prince obtained significant evidence of ESP when he applied this method to some of the nineteenth century experiments and also to the tests carried out by Upton Sinclair (1932/1960). But as Pratt (1937) points out, though Prince's method was "a step in the right direction", it nevertheless "depends for its validity upon having a fairly wide difference in ESP ability between the true subject and the (control)-subjects. It would fail in case there were either only a slight gift on the part of the true subject, or an equally marked ability shown by half or more of the (control)-subjects" (p.251).

The Method of Correct Matchings. The late Gaither Pratt (1937), reporting on some drawing-reproduction tests carried out by Dr. C. Hilton Rice, applied to the data (for the first time in parapsychology) a method widely used in psychological work, called the 'Method of Correct (or Forced) Matchings'. The technique requires no evaluation of the degree of target-response resemblances. Essentially, the set of n targets is displayed, together with its set of n responses, the drawings in each set being in randomized order; a 'blind' judge is required to match, in pairwise fashion, each response with its intended target, having nothing but the intrinsic similarities to work from; the response is either correctly matched with its own target (in which case it is called a 'hit', or 'correct matching') or it is matched with a target for which another response was intended (in which case it is a 'miss' or 'incorrect matching'); the total number of correct matchings may thus range from zero to n (with the exclusion of n-1). Chapman (1934) was the first to work out the exact probability distribution of obtaining r correct matches for n target-response pairs. For the case n = 10, using a single judge, Carington (1940a, p.70) gives the probability p of obtaining exactly r number of correct matches, and the probability p' of obtaining r or more correct matches. (For example, the method has yielded results significant at better than the 5% level if four or more correct matchings are made out of ten tries.) Indeed, Stevens (1939) showed that for a single matching, the most likely number of correct matches for a single percipient is unity, with variance same. If one uses k number of judges to match, independently, the same set

of n target-response pairs, then Chapman (1934) also gives the probability, for various n's and k's, of obtaining a given mean number of correct matches (this probability thus being really a test of the reliability of different judges obtaining the given number of correct matchings).* If, on the other hand, we have x number of independent sets of n target-response pairs (e.g. n responses produced by each of x percipients), then it is convenient to apply the Stevens Matching Formula again, from which it can be shown that as n becomes large, the expected total number of hits is equal to x, with variance the same. Scott (1972) provides a convenient table for finding the probabilities associated with obtaining h or more total hits from x sets of n target-response pairs.

This matching method has been used not only by Pratt (1937), but also by Carington (1940a) and by Targ & Puthoff (1974a); it was used on one occasion in the present series of experiments (Chapter 5). It has its advantages in that it is relatively quick and simple to apply, is useful if only one judge is available and one wishes to avoid jeopardizing the Assumption of Independence (see below), and if one wants a 'rough-and-ready' guide as to whether ESP is occurring strongly in the data. The method gives good results when, as Stuart (1942) points out, "the (target)-drawings are sufficiently discrete items, and the subjects' reproductions are clearly differentiated from one another and match the (targets) to a striking degree".

---

* This fact tells us nothing, however, concerning whether the mean number of correct matches obtained significantly exceeds chance expectation. Pratt (1937) unfortunately confused the two; he was, quite properly, taken to task by Ellson (1940), and gallantly admitted his error (Pratt, 1940).

(p.23). The disadvantage is its conservativeness: the relation of a given response to a given target is treated in evaluation as an 'all-or-none' item -- that is, as a 'hit' or a 'miss' -- with no intermediate degrees; sensitivity is blunted if two or more of the targets (or responses) tend to resemble each other, in form or concept, because it thereby becomes harder for the judge to discriminate between them.

The Method of Palpable Hits. In evolving a workable assessment technique, Carington (1940a) came to the conclusion that it was necessary to abandon (for the time being) his hope of giving full scope to possible resemblances of theme, topology, symbolism, etc., and concentrate upon what he called 'palpable' hits. He persuaded a 'blind' independent judge to compare each of the response-drawings produced by his percipients against each of the targets used in a series of experiments; the judge was to keep in mind the principle: "do this drawing and that original plainly and unmistakeably portray the same thing? If they do, give one mark; if they do not, give nothing" (p.80); the response-drawing had to "provide plain and unmistakeable evidence that the object portrayed in the original was prominently in the mind of the percipient when he made the drawing" (p.80). The number of hits so scored on all targets was then apportioned into that number of hits obtained on the percipient's own targets and that number of hits on targets not used for that percipient; these numbers can be statistically compared to see if there is a relatively larger number of hits on own targets than on controls (bearing in mind the caveats suggested by Stevens (1940) and assented to by Carington (1940c)). One of the useful

features of this method is that, by using both actual and control targets, it can pick up evidence of temporal displacement occurring within the former, as well as contemporaneous hits. The technique has been used by Carington (1940a, 1941), by Marsh (1958, 1962), and most recently by Mittenecker & Schulter (1978).

But again, a problem with this method is, as Stuart (1942) comments, that the evaluation

> "assumes that the evidential features of the (target)-drawing which are transmitted or perceived should be sufficiently apparent and clear-cut that they can be given an all-or-none-classification. Since this is not a necessary assumption regarding ESP, it is apparent that the method suffers from just that degree of crudity of evaluation." (p.22).

And again (Stuart, 1946):

> "If a responding subject makes what is practically a facsimile drawing of the stimulus object, it is easily identified as a hit. If the response reproduces the essential idea of the stimulus, it is also, conventionally, a hit. But what about such cases as the following: stimulus: Soldier, and response: gun; or stimulus: Cat, and response: dog? These are not hits, but neither can the familiar associations involved in such responses be considered convincing misses. In work with unselected subjects the number of clear-cut hits is so small that unwieldy masses of observations are necessary to get enough cases to provide reliable statistical material." (pp.21-22)

The Catalogue Method. Because the judging task in the Method of Palpable Hits can be extremely burdensome, Carington (1944a) devised a simplified method of scoring by compiling a Catalogue purporting to list the empirically-derived 'chance frequency' that any particular object will be drawn in an experiment where that object was not also a target. This 'Catalogue Method' was used in a number of studies (e.g. Carington, 1944a,b; Carington & Heywood, 1944), and with an American version of the Catalogue (Taves, 1945) in the studies by Taves et al. (1945) and by

Schmeidler & Allison (1948). Despite the fact that the method was devised with the help of no less a statistician than Ronald Fisher, parapsychologists eventually came to the conclusion that the Catalogue frequencies were not sufficiently stable to allow reliable data-evaluation. For discussion of the problems involved, see Stuart (1944), Marsh (1962, p.8), Thouless (1963, pp.139-141; 1972, p.9) and Thalbourne (1976, p.9).

The Preferential Matching Method.* In an attempt to combine the best features of the Method of Correct Matchings and the Method of Palpable Hits, Stuart (1942), with the aid of Greenwood (1943), set up a technique whereby all kinds of recognizable target-response similarities could be used to arrive at a score representing this correspondence. This 'Preferential Matching Method' is often thought of as "simply a device for allowing partial credit" in the awarding of hits (Burdick & Kelly, 1977, p.112). The key to the method is rank-ordering: it involves displaying to a judge an 'evaluation-group' consisting of a set of n randomized comparison stimuli, and inducing the judge to rank these stimuli against one master-stimulus in accordance with their degree of similarity to it; the master-stimulus may be a response-drawing, in which case the evaluation-group would consist of a set of n target-drawings (all the targets used to test one percipient, a subset of these, or the target belonging to that response plus a number of equivalent 'control' targets not used at all for the percipient); conversely the master-stimulus may be a target-drawing, in which case the evaluation-group would consist of a set of n response-drawings

* Much of the following material echoes that published as Thalbourne (1979b).

(all of the response-drawings produced by a given percipient, or just a subset of them); in either case, a rank-score of 1 is awarded to the comparison-stimulus adjudged to correspond most highly to the master, a score of 2 to that stimulus next highest in correspondence, and so on down to a score of n for the least similar.

In the present series of experiments, the evaluation-group always contained ten fixed comparison-stimuli: either all ten responses produced by a given percipient, these being rank-ordered against one target (Chapters 3 through 5); or, on one occasion (Chapter 6), the converse of this procedure, where all ten targets used for a given percipient were ranked against one response, for purposes of methodological comparison. Since the composition of the evaluation-group is fixed from one ranking to another, this application of a matching method is referred to technically as a 'closed deck' case.

The rank-ordering which the judge gives the ten drawings in the evaluation-group constitutes an ordinal sequence, and thus the location of each response relative to each other can be represented by means of rank-scores ranging from 1 to 10 (without ties). The set of response-drawings is rank-ordered against each of ten target-drawings, thus resulting in ten sets of rank-scores, each of which scores describes the relative 'amount' of target-response similarity apparent to common-sense judgement, and which are fairly simply to compare with chance hypotheses.

At this point it will be necessary to make mention of the so-called Assumption of Independence. In order to ensure

that successive rank-orderings of a given evaluation-group are statistically independent of one another, a given judge should rank-order a given evaluation-group on one occasion only: if he or she were to rank-order the same group a second time, even though against a master-stimulus different from the previous occasion, the two resulting rank-orders may or may not be independent of each other, regardless of how the judge has been instructed to behave. In particular, the main problem arises if a judge has any tendency to avoid ranking as first a member of the evaluation-group he or she had previously ranked as first. For instance, suppose the first target depicted a Car, and the evaluation-group indeed contained a car as one of the ten responses; a judge will surely give the car-response highest rank; but faced with the same evaluation-group a second time, and a different target (say a Wheel), the judge may tend to rank the car-response low regardless of any similarity to the Wheel-response, on the (erroneous) assumption that the car-target/response pair must have been the correct match, and that now the car-response can be safely eliminated from consideration in all future rank-orderings of that same evaluation-group. Thus may bias, or non-independence, be introduced between successive rank-orderings.

In order not to jeopardize this Assumption, the experimenter rejected outright any idea of using the subjects themselves to rank-order their own material (which would necessitate rank-ordering the same evaluation-group on ten occasions). It was thought that the problem could be circumvented by employing 'blind' independent judges to rank-order the response-sets -- one judge for each of the ten targets, thus requiring a

minimum of ten judges; this, then, was the procedure used in this thesis. However, Kennedy (1979) and Markwick (personal communication) have each suggested that some dependence between the ten rank-orderings might still be possible if these 'independent' judges showed any tendency to rank-order the evaluation-group in a similar sequence due to (say) common aesthetic preferences or 'order effects' (the tendency to favour (say) the middle numbers of the evaluation-group display over the end members.) Fortunately, the test of statistical significance used (see below) does not, it was recently discovered, depend upon any assumption of independence; having more than even just a single judge was thus, after all, not necessary! (Having a sole judge would not, however, have been feasible in practice, because there were upwards of 300 rank-orderings to perform per experiment, and it takes about an hour to do just fifteen!)

Details of the Judging Procedure. Regarding the selection of judges, virtually no qualifications were deemed necessary other than adequate eyesight, reasonable intelligence and enthusiasm in the face of a somewhat boring task. Nearly all judges tended to be University undergraduates, frequently female psychology students in their late teens; in this regard, then, they resembled most of the subjects who took part in the actual ESP test. All were unpaid volunteers.

Before undertaking the ranking-task, the would-be judge was given some general instructions as to what sorts of similarity could be expected between target and response: for the literature (e.g. Delin, 1977) reports that not only obvious, perceptual (or 'topological') resemblances are found but also less obvious conceptual, thematic or associational similarities. Each judge

therefore underwent a standardized 20-minute 'audio-visual' training program: for illustrative purposes, they were shown enlarged photographic reproductions of the targets and responses from Targ & Puthoff's (1974a) drawing-test with Geller (see Appendix 8); simultaneously, they heard a tape-recorded monologue in which the author explained all the necessary principles of the judging-procedure and pointed out all the target-response resemblances to be found in the Geller data (for a transcript of this monologue, see Appendix 9).

As a preliminary to the judging phase of an experiment, the ten responses comprising each evaluation-group would have been randomized, the random-order being different from one evaluation-group to another. Thus, since the response-sheets were numbered by code rather than by sequential order, there was, on this level at least, no possible means by which the judge could tell which response was the correct match for the target: he or she had to work purely on the basis of similarity.

After the instruction-period, the ten response-drawings of a given percipient were laid out, randomized, end to end on a long table. The judge then proceeded to rank-order each pair's ten response-drawings for degree of correspondence to one only of the targets that had been used for that pair. Note that while the judge was working thus, the experimenter (who was also the judging-supervisor) remained unobtrusively in the background, in another part of the room: he would generally be very occupied in recording the rank-scores of the responses previously ranked, in putting those responses back into their predetermined random-order (ready for the next judge), and in laying out the next evaluation-group ready for the present

judge to rank. This is mentioned to reduce the plausibility of the suggestion that the experimenter could have unconsciously given the judge tiny sensory cues as to which response was the correct match for the target, in that minority of cases where he remembered the correct pairing. Even so, it was eventually established as routine procedure that experimenter was out of the judging-room as much as possible, and that, rather than him laying out the next evaluation-group, the judge would be given the packet containing the response-drawings (in their randomized order) and told to take them out and lay them out himself for display. This aspect of the procedure could thus be better controlled if a blind assistant was able to be employed to interact with the judges; but since the judging-phase is extremely time-consuming and tedious, this possibility was out of the question in the present series of studies.

It may perhaps be useful to mention that though there are wide individual differences between judges in the amount of time they take to perform a single rank-ordering one may expect that on average approximately 15 rank-orderings (of different evaluation-groups) can be completed within one hour, after which the judge's energy and interest begin to flag (which is detrimental to the evaluation).

The resulting rank-scores awarded to the response-drawings were entered horizontally into a special target-by-response matrix -- one matrix per agent-percipient pair (see Appendix 10). Provided reasonable care is taken, the recording of these scores is fairly fool-proof, since errors are very obvious and thus self-correcting. (A special computer-program has been written by the author to check that the rank-scores 1 to 10 have all been entered into each row of the matrix.)

The Use of Multiple Independent Judges. An unfortunate aspect of the single-judge procedure (that is, the case where an evaluation-group is ranked against a given target by one person only) is that so much weight is given to the opinions of just one adjudicator. Some judges, while conscientious enough, will occasionally make evaluations that can only be described as being highly idiosyncratic. The author once had a judge who, when faced with a train-target, ranked a car-response as ninth most similar out of ten (i.e. as the second most dissimilar)! When later questioned, she confessed that until it had been pointed out to her, she had not thought of any conceptual or topological relationship between the car and train, whereas no fewer than five other judges (each blind and independent of one another) ranked the car-response as first in similarity!

Such deviance results in a reduction in both the validity and reliability of the rank-scores (different judges may award totally different scores). This is obviously the very antithesis of a trustworthy method of assessing target-response resemblance. Another curious property of the single-judge regime is that even if the judge achieves an extra-chance number of correct target-response matches, this may have been due to his own psi rather than that of the subjects who produced the responses! This is particularly a problem if the resemblances claimed to be present are 'in fact' very far-fetched and not at all apparent to inspection. (One is reminded of the 'Da Vinci phenomenon', in which imagery is stimulated by the patterns of flames in a fire.)

The author would like to suggest that at the price of an increased amount of experimenter-labour, a greater degree

of validity and reliability may be achieved by using a multiple-judging scheme. In the present series of studies, triple-judging was employed: the ten responses of a given subject were rank-ordered against one target, by three 'blind' and independent judges. Since there were ten targets used for each percipient, a minimum of 30 judges were therefore required for each experiment. ('Minimum', because unless all 30 judges rank-order all N evaluation-groups (N frequently exceeding 15) additional judges had to be brought in to be trained to rank-order the evaluation-groups left over.)*

Thus, for each evaluation-group vis-à-vis one target, there were three sets of rank-scores. The resultant dependency among the three judge's rank-orderings was treated by using the mean of the three rank-scores awarded to each target-response pairing. By using this 'average-rank score', we can thereby help to 'smooth out' occasional idiosyncratic rankings, and moreover reach a more acceptable 'consensus of opinion' about the target-response resemblances: for if we can say that three persons have each, independently, awarded similar rank-scores, then we are more justified in concluding that the scores reflect similarities which are 'really' there and about which most reasonable people would agree.

---

* It would have been desirable, though perhaps not necessary, for successive judges to encounter a given evaluation-group with its constituent response-drawings in different random orders, to minimize any order effects on the part of judges in the matching. Consistent order effects might give rise to spuriously high inter-judge concordance, and it would be an interesting methodological experiment to compare concordance levels for 'repeatedly-randomized' and 'non-repeatedly-randomized' orders of the same comparison stimuli. A later study, reported in Thalbourne (1980), evaluated the data with just the 'repeatedly randomized' order of comparison stimuli: a level of concordance was obtained (mean W = 0.66) which was the highest for any of the groups of data ever evaluated in the author's experiments! This obviously offers little support for the hypothesis that concordance is seriously inflated by using a 'non-repeatedly-randomized' order.

The appropriate statistic measuring the degree of agreement between the three judges is Kendall's W coefficient of concordance (Siegel, 1956): the value of this measure ranges from zero (no agreement at all) to unity (perfect agreement), and for three judges rank-ordering ten objects must be equal to or greater than 0.627 before we can conclude that the judges are applying basically the same criterion rather than simply rank-ordering randomly. In the author's experience, there is by no means perfect agreement (1.00) amongst multiple judges: though W can be as high as 0.96, it can be as low as 0.14, and averages out at about 0.63 (n = 1140). It can be shown that W is equal to an average Spearman correlation between all possible pairs of the K judges, and in this case a mean W of 0.63 indicates a mean rho of only +.44 for an evaluation-group containing ten objects. Though this figure is probably acceptable*, it does nevertheless indicate that it would be unwise to rely on a single judge's rank-ordering, at least as regards establishing of the 'objectivity' of the target-response similarities. Additional advantages of the multiple-judging regime are that (i) even when concordance is low, the result of averaging divergent rank-scores is average-ranks that converge upon mean chance expectation, thus (as it should in the circumstances) preventing spurious (or unreliable) evidence of extra-chance scoring; and (ii) if the concordance

---

* Any researcher keen to undertake drawing-tests in the future might be well advised to have a number of trials smaller than ten -- four, or perhaps six at the most: this would be less tiring for the agent and percipient, almost certainly easier for judges to evaluate, and therefore might lead to a higher degree of inter-judge concordance. There is also some evidence that most of the psi in ten-trial drawing-tests occurs during the first six trials (see later discussions of so-called 'serial-position effects').

is high, then this renders less plausible (even if it cannot totally eliminate) the suggestion that ESP could have been exercised by the three judges in the choices they made rather than by the percipients. Thus, the subjective opinions of a plurality of adjudicators should be pooled to form a reasonably objective consensus of opinion.

So much for the question of validity. What about the reliability of the average-rank scores? Since all subject-pairs were tested once only in this series of experiments, we have no data available concerning the correlation between the ESP-scores of the same subjects tested on different occasions. But data have been obtained regarding the correlation between the two average-rank scores yielded by having different judges evaluate the same data: we can call this 'judge-rejudge reliability'.

A mean W of 0.63 -- which is typical for the present data -- indicates an average Spearman correlation between all possible pairings of the three judges of about +.44. This suggests some overlap between their rank-orderings, but still a good deal of variability. The figure of +.44 represents an estimate of the judge-rejudge reliability on a single-judge regime. What is the corresponding figure for the triple-judging regime? What, in other words, is the stability of the average-rank scores? A small experiment was carried out in which some sets of response-drawings already evaluated using the triple-judging method were presented to entirely new trios of independent judges. The 160 target-response pairs re-evaluated in this way, thus had two average-rank-scores now associated with each, each purporting to measure the 'real' degree of target-response resemblance. The old scores correlated with the new to a value of r = +.71

($2p = 1 \times 10^{-10}$). Thus, though there is still some unreliability involved (50% of the variance remaining unaccounted for), this is a far better degree of stability than from one single-judge to another. The range of average-rank scores is in fact from 1.00 to 10.00, and therefore the range of possible difference-scores between initial ranking and re-ranking is 9. For the sample of 160, the average difference-score was -0.001 (indicating no systematic tendency for the re-ranks to be greater or lesser than the original ranks). Of all the difference-scores, none exceeded 4.66; 54% of the differences were less than 1.33, 79% less than 2.33; and 92% less than 3.00. Thus, results obtained from a triple-judging scheme should be reasonably generalizable from one judging-evaluation to another.

The Test of Significance for a Subject-Pair's Ten Average-Ranks.

Having transformed qualitative degree of target-response resemblance into a rank-score, it then becomes necessary to determine (1) what value of rank-score (i.e. what degree of similarity) may be expected to arise by chance, as well as (2) how likely is the deviation from chance of our observed rank-scores. The answer to the first question is that the average rank-score to be expected by chance over a large number of trials when N objects have been rank-ordered is $(N + 1)/2$, or in the case of N = 10, 5.50. For T number of trials, the expected sum of ranks is $(T \times 5.50)$, or for T = 10 (in the case of a single agent-percipient pair), 55. Had a single-judging scheme been used, we could have tested the significance of the sum-of-ranks by using the Exact Test described by Morris (1972) and extended by Solfvin, Kelly and Burdick (1978); if

the assumption of independence was thought to be in doubt, we could have used Scott's (1972) Permutation Method; or even, under appropriate circumstances, the so-called 'critical ratio', or z-score transformation (in which the deviation from chance of the obtained sum-of-ranks is divided by an estimate of the standard deviation of such sums-of-ranks: see Stuart (1942), Greenwood (1943), and Solfvin et al. (1978). However, the use of average-ranks derived from multiple-judging is very apt to result in fractional numbers; this, and the fact of less-than-perfect inter-judge concordance, leads to distributional properties of the average-ranks that are rather different from those of the single-judge rank-score distribution. Due to these differences, neither Morris' Exact test nor the Critical Ratio method can be used.* How, then, can the significance of the average-ranks be tested?

The procedure adopted in the present thesis was to use a computer-run Randomization Test.** The rationale behind this test is as follows. For each subject-pair, the process of triple-judging has yielded a ten-by-ten target-by-response matrix of

---

* According to Thalbourne (1979b), one could perhaps choose, from each triad of rank-scores awarded by the three judges the median rank-score and use this to compute a 'quasi-single-judge' sum-of-ranks. Betty Markwick (personal communication, 1980) has pointed out that in fact, owing to less-than-perfect inter-judge concordance, this procedure is not, after all, statistically permissible, because there would be a deficit in extreme rank-scores available.

** Betty Markwick (personal communication, 1980) has pointed out that in fact the so-called Greville Formulae (Greville, 1944) can also be applied to the present data: this method does not require the use of a computer, and seems to be only marginally less powerful than the Randomization Test.

average rank-scores. (See, for example, the partially-filled matrix in Appendix 10.) Targets are listed down the left-hand side in order of presentation to the agent, and responses are listed from left to right along the top, in order of production. Each of the 100 cells in this matrix contains an average rank-score which represents the degree of relative similarity obtaining between any given target and any given response -- in fact all possible pairings of targets and responses. The principal diagonal (i.e. that running from upper left- to the bottom right-hand corner) contains the ten cells representing the correct target-response pairings, and their associated average-ranks. The 90 off-diagonal entries represent the degree of resemblance between incorrect target-response pairings (e.g. first target with eighth response, etc.). These 90 scores are just as 'valid' as the ten diagonal scores, in the sense that they do represent objectively the degree of similarity between target and response; it is simply that these target-response resemblances have arisen purely by chance; or, at the very least, they are not the result of contemporaneous GESP (although they may of course have arisen as a result of displacement, but this would not be relevant to the argument).

Now on the null hypothesis, the correct, or contemporaneous target-response pairings likewise show only a degree of resemblance attributable to coincidence: on the null hypothesis, there is no reason to expect the values of the diagonal scores to be any different from those in the rest of the matrix. The Randomization Test examines the alternative hypothesis that there is something different about the ranks along the diagonal; it tests the hypothesis that the scores in the matrix as a whole are not distributed in a purely chance fashion. The method consists of a comparison between the ten diagonal ranks and the 100 ranks in the matrix as a whole. If contemporaneous GESP did not influence the production of the response-drawings, then the value of any desired diagonal statistic should not be "significantly different from" (i.e. "very rare compared with") the statistic obtained when any ten cells are randomly selected from the matrix (given that this random selection procedure is constrained by the same limitations as is the diagonal, namely that there can be no more than one cell chosen from a given column and a given row.*) The entire ten-by-ten grid, including the diagonal, should be merely a chance array of scores.

The number of possible sets of ten entries that can be made up out of 100 cells (when applying the row-column selection

---

* Thalbourne (1979b), in his extended description of the use of the Randomization Test, omitted to mention this vital constraint. Betty Markwick and James Kennedy (personal communication, 1979) have both pointed out that an 'unconstrained', or 'general' selection procedure leads to a spurious over-estimate of significance for the diagonal statistic; what is needed is the 'row-column' selection procedure, as described here.

procedure) is ten factorial (3,628,800). Though a modern computer could probably enumerate all the sets quite economically for matrices up to size seven-by-seven or eight-by-eight, the ten-by-ten matrix was deemed too large for this to be feasible. Therefore, a Fortran-language computer program was written, called "RANMAT" ("Randomization Test for a Matrix"), and which was instructed to take a subset of the ten factorial sets, performing all the random selections and comparisons required: ten entries were randomly selected (on the row-column constraints) from the matrix one thousand times, and the desired statistics were calculated each time before the entries were replaced back into the matrix for the next random selection. (It was assumed that 1,000 such sets would be adequately representative of the ten factorial possible sets.) The frequency with which a value of the desired statistic occurred in those 1,000 randomizations which was equal to or greater than the value of the actual diagonal statistic, was then tabulated and converted to a proportion. This figure could then be regarded as a quasi-exact probability -- a Randomization Test estimate of the exact probability that the value of the actual diagonal statistic would occur 'by chance'. In this way, then, a subject-pair's results could be shown to be so unlikely -- so rare on the null hypothesis -- as to be statistically significant.*

---

* The author would like to acknowledge his debt to Dr. Peter Delin of the University of Adelaide, who first suggested to him the idea of using a Randomization Test, and who worked out many of the statistical details involved. The author would also like to thank Mr. Bob Willson, who wrote the computer program RANMAT (a listing of which may be found in Appendix 22). It should be pointed out that the term 'Randomization Test' was used originally (e.g. Siegel, 1956, pp. 152-156) to refer to a procedure for determining exact probability by enumerating all possible combinations of the raw-data. In this sense, too, Scott's (1972) Permutation Method is also a Randomization Test. The test described in this thesis may well be more correctly termed a 'Monte Carlo Method', since only a subset of all possible

The most important advantage of a Randomization Test is its statistical power. The reason for this greater power is basically that it is distribution-free, in the sense of being free of any assumptions about the characteristics of the underlying distribution of scores, such as normality. RANMAT simply takes the scores in the matrix as it finds them, shuffles them about at random, and generates a probability distribution of its own. The fact that the Randomization Test is here run on a computer is due in part to its convenience but also because full evaluation of the exact probability would not be feasible.

A consequence of the fact that the Randomization Test probability is not exact but only an estimate, is that if the same matrix is analyzed several times, the probability-estimates will not all be exactly the same. What degree of variation do we find on repeated runs of RANMAT? How reliable are the p-values so obtained? For non-extreme values of the statistic, the difference in probabilities can be as much as 0.038, according to observation of the outcome of RANMAT; fortunately, for non-extreme statistics, the question of their being statistically significant is an academic one. However, the more extreme from mean chance expectation is the statistic, the smaller the variation obtained on a re-run of RANMAT. A mini-experiment was conducted in which each of four matrices was re-run at least ten times (total N = 50). All these matrices had initially yielded a significant diagonal statistic. The average range of differences in p-values turned out to be 0.005, with an s.d. of 0.002: hence, one can expect in general a change of

---

combinations are selected and used to generate an empirical distribution of the desired statistic. The use of this approach in psychical research seems to have originated with George Medhurst (see Barrington, 1973).

between 0.003 and 0.007 in p-values on repeated runs of RANMAT.

Choice of Statistics. The most obvious statistic to select is the pair's mean average-rank score (or 'mean rank' for short): this gives an indication of the average degree of resemblance between target and response over the ten trials of the ESP test. For example,a pair with a mean rank of 3.50 tended to produce response-drawings that were somewhat more similar to their own targets than would be expected on average (namely, 5.50). The Randomization Test will tell us whether 3.50, for that particular subject-pair, deviates so far from 5.50 that it is unlikely to have arisen by chance. Now because the values of the 100 average-rank scores comprising the ten-by-ten matrix will differ somewhat from pair to pair (in a fashion related to the degree of inter-judge concordance), the distribution of 'pseudo' mean ranks generated by the Randomization Test will vary as regards the size of its standard deviation. Thus, the exact probability of a given value of mean rank (say, 3.50) will likewise vary from matrix to matrix: if the s.d. is small, the value of mean rank will tend to be more significant (i.e. more extreme) than when the s.d. is large. Hence, the probability of a given mean rank must always be calculated de novo, from its parent matrix.

Nevertheless, the author has to date made 230 runs of the RANMAT program: since, on each run, the computer was instructed also to print out the values of mean rank at various percentile points of the distribution of Randomization Test 'pseudo' means, empirical observation gives us some idea of the likely values of mean rank required at the 'cut-off points' for the attainment of statistical significance. For the 'above

chance' end of the distribution (i.e. means lower than 5.50), it turns out that the obtained mean rank must on average be 4.12 to reach the 5% level of significance (two-tailed), though it may on occasion have to be as low as 3.83 or as high as 4.50. Similarly, for the 'below chance' end of the distribution (i.e. means above 5.50), the cut-off point is about 6.86, though again in any given case it may be as low as 6.57 or as high as 7.13. Thus, on average, mean rank must deviate from 5.50 to a value of about 1.37 units in order to reach statistical significance.

Mention of below-chance means prompts the note that the distribution of mean ranks is approximately normal and is symmetrical about 5.50, with the result that a mean of (say) 6.86 can be as rare an event, statistically, as 4.12 (which are the two average cut-offs for two-tailed significance at the 5% level). A mean of 6.86 indicates that a pair's pictures tended to be ranked less highly (i.e. to be more dissimilar) than one would expect by chance. Over ten trials, one expects that due to the operation of chance, a least a few responses will display some tenuous resemblance to their targets. If the responses consistently lack even this small amount of similarity, then this is a very remarkable and statistically rare phenomenon. Significant absence of similarity between target and response seems to be the picture-guessing analogue to significant negative scoring in card-guessing tests, when far fewer cards are correctly called than would be expected by chance. Systematic absence of similarity to the target-drawing presupposes knowledge of the target in some fashion: because of the way chance operates, the percipient cannot consistently avoid giving likenesses

to the targets, unless they have some sort of knowledge regarding what likenesses to avoid. Significantly below-chance scoring is known as 'psi-missing': the fact of significance allows us to postulate the existence of some cause for the low scoring, and parapsychologists call this cause 'psi'. Of course, like previous methods of assessment, the Randomization Test can easily cope with the psi-missing situation.

Psi-missing and psi-hitting are considered not to differ in their evidential value: Rao (1974), for example, points out that there is no good sense in which the hitter can be said to "have more ESP" than the misser: psi may express itself 'bidirectionally'. Many are the postulated causes of psi-missing. (Relevant discussions may be found in Rao (1965), Louisa Rhine (1967), J.B. Rhine (1969), and Child & Levi (1979).) Some of these parapsychologists have suggested that psi-missing is analogous to psychoanalytic repression or to other more mundane lapses of memory. This 'blockage' model may well apply to card-guessing situations: blockage of the correct response will mean that the percipient calls a symbol from the remaining four (incorrect) target-possibilities, and if this tendency persists it will result in an accumulation of misses that amount to a significantly low number of correct guesses. However, the present writer feels that this blockage model cannot apply to free-response situations: if the correct target (say 'Tree') were blocked from consciousness, and the percipient were left to guess from all the non-Tree possibilities, he could presumably still score above chance: for the percipient could as easily respond with 'leaf' or perhaps even 'umbrella', and despite the fact that these responses come from the set of all 'non-

Tree' responses, they would still be ranked fairly high in similarity to the Tree-target. For blockage to occur in such a way as to lead to significant negative scoring, the whole network of Tree-like responses would have to be supressed, which is to say, all conventional associations to Tree (such as 'flower', even 'dog'!) and at the same time all graphic responses that had some topological, or formal, similarity to the shape of the Tree (such as, perhaps, an umbrella). Thus, to achieve a significantly negative score, the response must not simply be a non-Tree, but rather something utterly un-Tree-like -- diametrically opposite a Tree in form and concept (e.g. perhaps, a window). The author therefore thinks it more parsimonious to drive an analogy between free-response psi-missing and psychoanalytic reaction-formation, in which the response conveys information directly opposite to what is known 'deep-down'. In fact, the suggestion may be made that at least some cases of psi-missing are instances of ESP being used in such a way as to avoid giving the appearance of 'success' (which is to say, above-chance scoring, or psi-hitting): motivation and need may be as relevant to psi-missing as they seem to be in the case of psi-hitting (Murphy, 1943).

But the phenomenon of psi-missing is further relevant for the choice of statistics. For since psi is so erratic and difficult to control, a subject-pair could conceivably psi-hit at one time in the experiment and psi-miss at another, resulting in a nonsignificant mean rank. For example, if five of their responses are so good that they receive average-rank scores of 1.00, and five so low in similarity to their targets

that their scores are 10.00, the resulting mean is 5.50, which is "at chance": the high scores would be cancelled out by the low scores. (This effect would be analogous to the 'cancellation effect' observed by Stanford & Brier (1968) within 'runs' consisting of 25 consecutive calls of ESP cards.) Yet ex hypothesi, ESP would still be occurring, albeit manifesting in two modes -- hitting and missing. Thus, the mean rank may disguise the fact that the individual scores are very extreme and the variation between them very great -- so great in fact as to be a very rare statistical event. Significant fluctuation around mean chance expectation may therefore indicate the dual operation of psi-hitting and -missing. Accordingly, the Randomization Test was programmed to provide information concerning an ad hoc statistic called the 'mean absolute-deviation' score: for each individual average-rank score, an 'absolute-deviation score' was calculated by subtracting the former from 5.50 and ignoring the sign of the resultant difference. These ten absolute deviations were then themselves averaged to obtain a mean absolute-deviation for each percipient. Whereas a subject's mean rank may indicate a tendency to have extreme scores in one direction, the mean absolute deviation measures the tendency to get extreme scores in both directions, regardless of whether they are high or low. (Neither the individual absolute-deviation scores, nor their means, have any tendency to correlate with individual average-rank scores or mean ranks. The judge-rejudge reliability for the individual absolute-deviation scores seems to be about +.43 ($N = 160$, $2p < 0.001$).) The beauty of the Randomization

Test in this context is that we can calculate the significance of any given mean absolute-deviation without any need whatsoever to know the population parameters of this score. (Empirical observation of repeated runs of RANMAT does, however, suggest that the mean chance expected score is in this case between 1.90 and 1.93.) Since also it would be rather difficult to explain the modus operandi of significantly low mean absolute-deviations (scores too close to 5.50!), it was therefore decided to examine only significantly large values of this score, thereby making the statistical hypothesis one-tailed rather than two-tailed as in the case of mean rank.*

Analogous arguments can be applied to the grand mean rank of a whole group of agent-percipient pairs: though the overall grand mean rank may well not differ significantly from chance, their grand mean absolute-deviation may do so. Also, the subject-pairs as a body may psi-hit on one trial and psi-miss on another, creating such a large degree of score-variation from trial to trial that it is significant. This would then constitute a 'serial-position' effect, i.e., an ESP effect in which the level of group-scoring on a particular trial is related to the position that that trial holds in the series of ten. When graphed, such patterns of scoring frequently reveal shapes that conform significantly to some simple mathematical function, such as a linear, quadratic or cubic trend: for example, in the 'Decline Effect', scores start high and decline to a point

---

* The mean absolute-deviation score is similar to, but not identical with the variance. The approach described here resembles the 'Symmetric Weighting Scheme' discussed by Solfvin et al. (1978).

at or below chance level; again, the term 'salience' refers to a U-shaped curve, in which the subjects start with high scores, progressively decline, but then pick up again on the last few trials. "The data so resemble other performance curves at repetitive tests -- the midafternoon factory letdown, for instance -- that they are usually interpreted as evidence of mood change, with initial interest giving way to increasing boredom." (Schmeidler, 1969, pp.17-18).

So, in view of the frequency with which ESP effects of this subtle kind have been reported, it was planned to analyze the group-scoring levels over time (i.e. over the ten picture-guessing trials). This can easily be done by means of one-way Repeated Measures Analysis of Variance. Tests for trend may also be carried out using the method of orthogonal polynomial coefficients.*

But before leaving the topic of data-analysis, mention must be made of one last score-type which was used in this present series of experiments, known as the 'hit-score'. To calculate this score, the ten average-rank scores obtained by a given subject-pair were each classified as denoting a 'hit' or a 'miss' depending upon whether the average-rank score was above chance (i.e. less than 5.50) or below chance (i.e. greater than 5.50). (No average-ranks were exactly 5.50, by virtue of the process of averaging the individual rank-scores awarded by three judges.) In other words, the scale of average-rank-scores, which ranges from 1.00 to 10.00, was turned into

---

* Some purists may object to the use of parametric statistics with data which are strictly speaking only ordinal in level-of-measurement. The present author takes as his imprimatur the statement by Solfvin et al. (1978,p.105) to the effect that: "Because of the extreme rapidity with which the ordinal weighting scheme ... converges to normality, it is appropriate to analyze group differences using t tests, analysis of variance, or other parametric techniques."

a dichotomy around the theoretical mid-point. Total number of hits per pair could then be readily found by addition, this score ranging from zero (no hits, all misses) to 10 (all hits, no misses). Individual hit-scores correlate about -.84 with their parent average-rank score, and there is a similar degree of correlation between the mean rank and the total-hits score. Thus, while the two score-types are not clones of each other, they are nevertheless close siblings. The judge-rejudge reliability of the individual hit-scores is approximately +.46.

The reasons why this score-type was used were that, firstly, it has a much more intuitively obvious interpretation (high scores indicating more similarity than do low scores, which is the reverse of the situation with mean ranks); secondly, the significance of a given total-hits score can be readily tested using the Binomial Test (with $P = \frac{1}{2}$); and finally, it provides a 'back-up' for the mean rank score: when the mean rank score yields only marginal significance, the hits-score will frequently reach significance, and vice versa, thus co-operatively providing a more sensitive test of the presence of ESP.

CHAPTER 3

The First Sheep-Goat Experiment

CHAPTER 3

"Before Him will be gathered all the nations, and He will separate them one from another as a shepherd separates the sheep from the goats ....."

Matthew XXV, 32

Introduction

Most of the experiments to be reported in this and subsequent chapters owe their inspiration to the first parapsychological study ever attempted by the writer -- a piece of research which was carried out as a part of the requirements for the Honours Degree in Psychology at the University of Adelaide (see Thalbourne, 1976). To understand the rationale underlying the doctoral research, it will therefore be necessary to back-track a little and describe this inaugural experiment in some detail.

The study had been designed in such a way that it would yield data of interest both to students of normal and paranormal psychology. Its parapsychological interest lay in the fact that an attempt was made to investigate what may be termed 'the closeness hypothesis', that is, the hypothesis (and widely held belief) that telepathy is more likely to occur between two human beings who share a close emotional relationship than between persons who do not have such a relationship. This popular notion was put to empirical test by examining the ESP-performance of two different groups of dyads, in the context of a ten-trial picture-guessing experiment such as has been described in Chapter 2: the experimental group consisted of 18 pairs of people who had said, via a brief self-report questionnaire (see Appendix 2), that they shared "a very special relationship", such that they felt "extremely close to one another on a personal and emotional level"; these people were termed 'close-relaters'. The control group really only needed to consist of pairs of people who did not share such a relationship with one

another; but in order to add psychological interest as well, the persons in this group were those whom the writer has dubbed 'non-close-relaters' -- that small minority of individuals who say, to the above-mentioned self-report questionnaire, that they have never in their lifetime been involved in a close personal and emotional relationship with anyone (even with a parent); it seems scarcely necessary to point out that randomly matched pairs of non-close-relaters would be suitable for the control group: inasmuch as they did not have a close relationship with anyone, they obviously would not have such a relationship with their experimental partner. Since such subjects are not easy to find, the control group ended up with a total of only 13 pairs altogether.

Now in addition to a free-response test of their ESP, these 62 subjects were administered a number of standard questionnaire tests (about which more below): this was in order to examine the purely psychological hypothesis that compared with close-relaters, non-close-relaters are less intelligent in social situations and/or more inclined to be introverted, and that this in some sense explains why they have not hitherto formed a very close attachment with anyone. (Parenthetically, it may be said that non-close-relaters in fact proved to be no less capable in social cognition, but that they were more lacking in social interactive skills, as well as being more shy and introverted. This finding is of relevance to the parapsychological aspect of the study, inasmuch as it means that the experimental and control groups differed not only on the 'closeness of relationship' variable, but also on various personality dimensions, (notably extraversion), which thus represent somewhat confounding variables.)

As for the ESP-test itself, three measures of psi-performance were employed: two of these (namely mean rank, and total number of hits) looked at overall level of psi-hitting or psi-missing, whereas the third measure (namely, mean absolute-deviation) examined the 'variance' of the trial-scores, that is, the tendency to oscillate between psi-hitting and psi-missing within

TABLE 3.1

195 Correlations Between Psychological Variables and Number of Hits out of 10, Closeness Study

	Agents (n=31)	CR A (n=18)	NCR A (n = 13)	Percipients (n = 31)	CR P (n=18)	NCR P (n=13)
Expression Grouping	-.31	-.41	-.28	-.12	-.15	+.04
Missing Cartoons	-.23	-.40	-.03	-.15	-.06	-.15
Picture Exchange	-.14	-.10	-.24	-.24	-.20	-.18
Social Intelligence	-.36*	-.46	-.29	-.20	-.17	-.11
Affectothymia	+.14	+.23	+.01	+.49**	+.50*	+.28
Parmia	-.08	-.11	-.17	+.24	+.29	-.03
Self-sufficiency	-.33	-.39	-.17	-.32	-.42	0.00
Fdm from Withdrawing Tendencies	-.02	0.00	-.09	+.29	+.31	+.24
Social Skills	-.13	-.38	+.06	+.33	+.16	+.46
Social Introversion	+.05	0.00	+.22	-.40*	-.40	-.22
Relaxation	-.26	-.41	+.15	+.04	+.08	+.11
Interest	+.26	+.20	+.27	-.21	-.07	-.41
Agitation	-.05	-.16	+.06	+.08	+.11	-.20
Uncertainty	-.11	-.01	-.48	+.07	+.10	+.06
Enthusiasm	+.16	+.15	+.29	+.01	+.17	-.14
Attitude-scale 3	+.33	+.13	+.57*	+.52**	+.24	+.80***
" 4	+.29	+.09	+.55	+.50**	+.24	+.77**
" 5	+.21	-.01	+.42	+.49**	+.27	+.70**
Belief in ESP	+.05	+.11	+.06	+.40*	-.22	+.63*
Personal experience of ESP	+.18	+.21	0.00	+.39*	+.06	+.55
Believeself to be psychic	-.04	-.31	+.09	+.51**	+.32	+.74**
Have had a hunch	+.02	-.20	+.21	+.17	-.25	+.36
" " premonition	0.00	-.16	+.17	+.35	+.15	+.41
" " Precognitive dream	+.46**	+.43	+.30	+.51**	+.32	+.70**
" " vision	+.09	+.40	-.31	+.22	+.13	-----
Believe in life after death	+.38*	+.19	+.62*	-.07	-.22	+.48
" " spirit contact	+.41*	+.33	+.45	+.34	+.14	+.40
Telepathic experience with partner	+.23	+.07	-----	+.41*	+.46	-----
" " with non-partner	+.33	-.03	+.66*	+.54**	+.44	+.47
Have had telepathic experience	+.35*	-.05	+.66*	+.61***	+.61**	+.47
Estimated probability of ESP	+.34	+.26	+.25	+.42*	+.19	+.58*
Believe ESP occurred	+.07	-.14	+.51	+.13	+.04	+.20
" was inhibiting factor	-.16	-.25	-.06	+.35	+.39	+.34

Note: 'CR A' = 'Close-relater Agents'
'NCR A" = 'Non-Close-relater Agents'
'CR P' = 'Close-relater Percipients'
'NCR P' = 'Non-Close-relater Percipients'

* : $2p \leqslant 0.05$
** : $2p \leqslant 0.01$
*** : $2p \leqslant 0.001$

the one ten-trial session. The grand mean scores of the close-relaters were expected, in all three cases, to be higher than those of the non-close-relaters. Though the difference between the group-means was in the predicted direction for two out of three of the score-types, only the hits-measure yielded anything remotely resembling a significant difference, and this by a one-tailed non-parametric test: the close-relaters scored exactly at chance, with a mean of 5 hits out of a possible ten; the non-close-relaters scored significantly below chance, with a mean of 4.23 hits ($t = 2.25$, 12 df, $2p = 0.044$); and the difference was just significant by Mann-Whitney U-test ($z = 1.70$, $1p = 0.045$). Considering the total number of statistical analyses carried out on the data, this significance should not be regarded as more than slightly suggestive; certainly, closeness of relationship did not produce any dramatic difference between the scores of the two groups.

Nevertheless, the question arose in the writer's mind: "Was there any systematic relationship between the scores on the ESP-test and psychological variables other than closeness of relationship? Was there any evidence that those subjects who obtained high ESP-scores differed psychologically in some consistent way from those whose scores were low?" It seemed both possible and desirable to investigate this question, for as a result of the social-psychological part of the experiment, there were data from a whole battery of psychological tests, as well as from measures of belief in and experience of psi, and of mood as experienced during the actual ESP-test. In all, there was a grand total of 33 psychological variables, and they are listed down the left hand side of Table 3.1 In the following paragraphs are brief descriptions of these variables.

Three of O'Sullivan & Guilford's (1966) Six Factor Tests of Social Intelligence were used, namely, 'Expression Grouping', 'Missing Cartoons' and 'Picture Exchange'. These tests are said to measure skill in social cognition, which

includes such things as the ability to perceive and correctly interpret facial expressions, postures and gestures, and complex social interactions involving two or more persons. (A fourth variable, labelled simply 'Social Intelligence', is a composite score representing the sum of the scores obtained on each of the three tests.)

The next group of variables comprise six questionnaire-scales taken from three established personality tests. From the Cattell Sixteen Personality Factor (16PF) Questionnaire were taken Factor A (Affectothymia, or outgoingness), Factor H (Parmia, or social boldness) and Factor Q_2 (Self-sufficiency); all these primary factors make large contributions to Cattell's second-order factor Introversion. From the California Test of Personality (Thorpe et al., 1953) were taken two scales, namely, Freedom from Withdrawing Tendencies, and Social Skills (the latter measuring interactive rather than cognitive skills). And finally, the Social Introversion Scale was used — a questionnaire derived by Drake (1946) from the Minnesota Multiphasic Personality Inventory (MMPI).

The third major group of variables in Table 3.1 refers to five rating-scales filled out by each subject immediately following the telepathy test (see Appendix 7). These scales enabled the participant to describe their mood and emotional state during the picture-guessing test. Two questions were also asked regarding their opinion as to whether or not ESP had indeed taken place, and also whether they believed there had been any psi-inhibitory factor operative during the test; these two items (Appendix 7, top of page), are to be found right at the bottom of the Table.

Finally, there was a short attitude-to-psi inventory, which consisted of 13 questions concerning the subject's belief in, and experience of, various sorts of putative psychic phenomena. Eleven of these questions may be found in Appendix 3. The two remaining items were: "I believe I have had at least one experience of telepathy between myself and the person who is my

partner in this experiment", and "I believe I have had at least one experience of telepathy between myself and someone other than the person who is my partner in this experiment". Additionally, the answers to these 13 questions were combined in various ways to form several versions of an attitude-scale: Scale 3 comprises all the 11 questions found in Appendix 3; Scale 4 is identical to Scale 3 except that it omits the item 'Estimated probability of ESP occurring'; and Scale 5 is composed of the five items that correlated most highly with Scale 3, namely, Personal Experience of ESP, Hunch, Premonition, Precognitive Dream and Experience of Telepathy. Attitude-scale 3, which is the most complete, is the one which will be referred to most frequently in the text to follow. All three scales do, nevertheless, intercorrelate at values exceeding r = +.90, and thus they can each be taken, for all intents and purposes, as measuring the same generalized attitude-to-psi variable.

As regards the measures of ESP-performance, brief mention of them has already been made above (p.87): the three primary measures were mean rank, total number of hits, and mean absolute-deviation. It should be pointed out that, in addition, secondary scores were derived from the primary measures, by classifying each given score as lying either above chance, at chance, or below. Thus, there were in toto six ESP-measures for each subject, though they all intercorrelate so very highly that it is most useful to think of them as essentially two distinct measures, one looking at overall psi-hitting or psi-missing, the other at 'variance'. To simplify presentation of the data, Table 3.1 contains the relationships between the psychological predictor variables and just one of those six ESP-measures, namely, that score which represents the total number of hits obtainedbut of a possible ten. (There are thus five other tables that could be shown, but three of these would give results very similar to those in Table 3.1, while the two tables concerning the variance scores yielded so little statistical significance that it is not worth

presenting them in detail.) Concerning this 'hits' variable, it is to be remembered that there were ten trials altogether in the drawing experiment, and the response-drawing produced on each trial by the percipient may be classified as a 'hit' or a 'miss', depending on the degree of resemblance it displays to its target. Thus, the range of possible total scores is from 0 (no hits, all misses) to 10 (all hits, no misses), and mean chance expectation is 5 (that is, five hits and five misses).

The statistical technique employed to examine the relationship between the ESP-measures on the one hand, and the 33 psychological variables on the other hand, was bivariate correlation analysis: Pearson or Spearman coefficients were computed, depending upon the level-of-measurement of each pair of variables. (In Table 3.1, the correlations from 'Expression Grouping' down to and including Attitude-scale 5 — that is to say, the top half of the table — are all Pearson, whereas the remainder of the correlations are Spearman, a non-parametric correlation being necessary because the attitude-questions each constitute only an ordinal level of measurement.) In each case, a two-tailed test of significance was applied, and alpha was set at 5%.

At this juncture, it is necessary to mention a slight methodological difficulty in regard to the calculation of these correlations. Since the psi-test was a test of general extrasensory perception (GESP), any given score on an ESP-measure was, strictly speaking, obtained by the pair of persons, that is, by both agent and percipient working as a team. It seems to the writer doubtful whether one can say a priori that the score belongs only to the percipient, and not to the agent as well: the agent may be good at 'transmitting' the target, or the percipient good at clairvoyance, or both may be the case. In other words, it seems to the writer quite valid to say that the agent and percipient both have a score, namely, the score that the pair obtained. Now obviously, if both parties are allotted the same score, then the scores of agent and percipient are not independent of each other. Therefore,

we cannot lump the ESP-scores of agents and percipients into a single group and then perform our desired computations: if we did, we would be violating the statistical assumption that all the scores are independent of each other; we would be committing the statistical sin of 'artificially inflating the N'. But there seems no reason why we should not consider the agents as one distinct group, and the percipients as a separate group, and then, perform the correlations for each group separately.

Moreover, it is possible and permissible to break these two major groupings-by-role down even further, on the basis of whether the subjects came from the experimental (i.e. close) group or from the control (i.e. non-close) group: thus, for example, the 31 agents can be divided into two groups, one containing the 18 agents who were close-relaters, and the other comprising the 13 agents who were non-close-relaters; and likewise for the 31 percipients, as illustrated schematically in Table 3.2. This procedure has

Table 3.2 Schematic representation of the breakdown, into six groupings, of the 62 Ss in the Closeness Study.

		Agents (n=31)	Percipients (n=31)
18 Close-relater pairs	Pair 1	ESP-score x	x
	" 2	" y	y
	" 3	" z	z
	⋮	⋮	⋮
	" 18	" a	a
13 Non-close-relater pairs	" 19	" b	b
	" 20	" c	c
	⋮	⋮	⋮
	" 31	" d	d

the effect of controlling for the closeness variable. Thus, there are six major groupings of the 62 subjects: agents, close-relater agents, and non-close-relater agents; percipients, close-relater percipients and non-close-relater percipients. The six ESP-scores obtained by the members of each of these six groupings, were correlated with the 33 psychological variables, giving a theoretical total of 1,188 coefficients; however, from that number must be subtracted 18 coefficients which could not be computed owing to lack of variance, giving us 1,170 valid correlations.

The question immediately arises, of course: "How many coefficients would one expect to be significant, just by chance, given that we were performing a very large number of correlations?" Theoretically, one would expect that out of a thousand correlations between independent variables, 50 of them, or 5%, would be significant at the 0.05 level or better. As it happened, the total number of significant correlations which the present analysis yielded was a staggering 145, or just over 12%! Superficially, this seems well in excess of the number to be expected by chance. But a complication enters here, and that is the fact that not all of the 1,170 correlations were between variables that were independent of each other. For example, the three attitude-scales correlate with each other to a very high degree, and therefore, if one scale correlates significantly with ESP-score, then it is virtually guaranteed that the other scales will likewise correlate significantly with ESP-score; thus, in the case of the 13 non-close-relater percipients (see Table 3.1, column six), we have not three significant correlations ranging from +.70 to +.80, but really only one. So the vexing question remains: given six not-entirely-independent measures of ESP, and 33 not-entirely-independent psychological dimensions, just how many correlations would one expect to be significant merely as the result of chance? Is it the case that the figure of 12% genuinely exceeds what would be expected by coincidence? Given 1,170 correlations, was there really an ESP effect, or was it a mere statistical artifact, inflated by the

interrelatedness of the predictor variables? These are very crucial issues, and they will be addressed more formally in the next chapter. For the present, it is sufficient to point out that the experimenter's intuitive judgement at the time was that the data did display a genuine indication of paranormal cognition, and that his subsequent research was predicated on that assumption. There seemed to be just too many extremely high, extremely significant correlations, for it to be merely a chance effect.

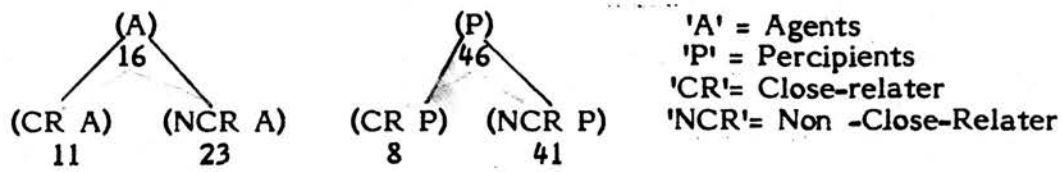
It was thus assumed that although some of the 145 significant correlations must have been chance-produced, by no means all of them were, and the question was, which ones were they? Now it is certainly the case that, prior to the computation of the thousand-odd coefficients, the writer did not go about setting up hypotheses in any formal manner. Nevertheless, common sense and our heritage of parapsychological research did give rise to various expectations about the possible outcome of this huge correlational analysis. For example, telepathy could be conceptualized as being a sort of social skill, and so it might be expected that social intelligence would correlate positively with ESP-score. Similarly, one might expect measures of extraversion to correlate positively with psi (see, for example, Eysenck, 1967; Palmer, 1978, pp. 132-133), and the same goes for such moods as relaxation and enthusiasm. And of course all the research on the so-called 'sheep-goat variable' (that is, belief in ESP, either in the abstract or with respect to one's own psychic ability: see Palmer, 1971, 1972; 1978, pp. 153-160) would lead us to expect a positive correlation between ESP-score and certain of the attitude-questions.

So, then, to the results, in Table 3.1. It can be seen, for example, that the mood-scales were quite unsuccessful in predicting the direction of ESP-scoring. Also, the relationships with the social intelligence variables are, contrary to expectation, negative, and rather consistently so (which might suggest that perhaps psi operates to compensate for a lack of the usual

sensory-motor social savoir-faire). Again, there is a reasonably consistent tendency (more pronounced in the percipients than in the agents) for high ESP-scores to be associated with extraversion, and low scores with introversion; only three of the 36 correlations are significant, but none of the three favour the introverts as high scorers, which is a result that agrees with a good deal of extraversion/ESP research.

But it is with the attitude-scales and their component questions that the vast bulk of the significant correlations in Table 3.1 are to be found. Though attitude-variables comprise only 54% of all correlations performed, they account for 86% of all the significant correlations. It is noteworthy, for example, that in the case of the 31 percipients (column four), all but one of the attitude questions correlate positively with ESP-score, eight of the 13 doing so significantly. Indeed, none of the significant correlations are in the negative direction. The overall impression is thus one of striking, albeit post hoc, confirmation of a 'sheep-goat effect', that is, a relationship such that 'sheep' (subjects who believed in psi and reported experiencing a number of types of cognitive psi) scored consistently higher on the experimental ESP task than did 'goats' (that is, subjects who tended not to believe in ESP and who reported few or no experiences of a paranormal nature). In the case of the 13 non-close-relater percipients (column six), the most complete version of the attitude-scale (namely, number 3), correlated with the total number of hits to the tune of +.80 ($2p = 0.001$). Despite the small number of subjects involved, and the large number of analyses performed, such a correlation appears to be astonishingly high, and suggests a high degree of predictive power: given a subject's score on this attitude-scale, one could fairly accurately predict their score on the ESP test, high scorers on the scale scoring above chance, low scale-scorers correspondingly scoring below.

Now given that there seemed to be some sort of ESP effect, the next question to ask was whether the effect occurred equally over all six of the subject-groupings. If we simply take head-counts of the total number of significant correlations for each of the six groupings, we obtain the results displayed below:



Note that for each group the same total number of correlations was calculated (namely, 195), and that therefore the figures given above are directly comparable. From this admittedly crude presentation, one could hazard a guess that the psychological make-up of the percipients was more strongly related to ESP-scoring than was that of the agents: percipients have nearly three times the number of significant correlations as do agents (46 versus 16). It can also be seen that the effect of controlling for closeness was a very drastic one: looking at the agents, the non-close-relaters obtained more than double the number for close-relater agents (23 versus 11); and for the percipients, the non-close-relaters obtained more than five times the number as did close-relaters (41 versus 8). In other words, the correlations were most likely to be significant when the subject was a percipient, and in particular when he or she was a non-close-relater percipient.

Why it should be that controlling for the closeness variable had such a dramatic effect, was certainly rather puzzling. Even so, the obvious empirical suggestion for a future research project was that subjects should be chosen on the basis of their being non-close-relaters, and their scores on the attitude-scale should then correlate positively with their ESP-scores. But this particular suggested line of research was rejected, for several reasons: (1) firstly, no plausible theoretical reason could be thought of why the correlations should hold for non-close-relaters but not for close-relaters; (2) secondly, on purely practical grounds, it would be much more difficult to do

an experiment consisting only of non-close-relaters: they are rather rare people (estimates ranging from between 9% and 16% of a psychology-student population), and they are not easy to persuade to be in ESP experiments, because they tend both to be very shy and also generally skeptical regarding psi; (3) but thirdly, and most importantly, was the fact that while it was clear that controlling for closeness did have a marked effect on the size of the sheep-goat correlation, it did not seem to affect the direction: the correlations for the percipients as a combined group of 31 were still quite significant. For instance, attitude-scale 3 (which shall hereafter be referred to simply as the 'sheep-goat scale') correlated +.52 with total number of hits (Table 3.1, column four), the two-tailed probability of which is 0.003. It seemed to be the case that even without controlling for closeness, the correlation would still hold, so why go to all the bother of trying to get specifically non-close-relaters? It was therefore decided to carry out a follow-up drawing-reproduction experiment, in an attempt to replicate this sheep-goat effect. It was reasoned that if subjects were selected who were extreme scorers on the attitude-scale (that is, either extreme sheep or extreme goats), then if the correlation were replicable, these subjects would accordingly be extreme scorers on the ESP test, that is, either psi-hitters or psi-missers, respectively. The remaining part of this chapter describes this experiment in detail.

METHOD

Subject Selection. In March 1977, the writer was given the opportunity to administer the ten-item sheep-goat scale* (plus the 'closeness of relationship questionnaire' and the six personality scales mentioned above [p. 90] in regard to the correlation analysis) to a large group of people, namely, 235 First-Year Psychology students at the University of Adelaide (South Australia). One of the aims of this mass-testing was to select out very

* That is, the ten items in Appendix 3, minus the Estimated Probability item.

extreme scorers on the sheep-goat scale. The plan was to test the ESP-performance of 15 pairs of extreme sheep and an equivalent number of extreme goats. The frequency distribution of the 235 scale-scores was thus obtained, and those 30 subjects scoring lowest were chosen as goats, while those 30 scoring highest were to be conscripted into the sheep group. The possible range of scores for this ten-item scale is 0-20, and in practice it turned out that the sheep were those persons who scored 14 or above, while the goats were those who scored 5 or below. Since the mean scale-score for the sample of 235 was 9.3, with an s.d. of 4.5, this meant in effect that any subject was selected whose score lay more than one standard deviation from the mean.

Characteristics of the Selected Sample. As it happened, only 13 pairs of goats could be recruited, largely on account of the difficulty of persuading extreme 'skeptics' to participate in a study of a phenomenon whose existence they doubt. By contrast, it was not at all hard to find willing and enthusiastic sheep, and in fact 17 pairs were tested. There was thus a total of 30 pairs. Of these 60 subjects, 27 were males, 33 females, fairly balanced across the two groups. Age ranged from 17-43 years, heavily skewed towards the younger end, median age being 18. Arts and science faculties were represented about equally. Most of the subjects were single, though ten were married and one divorced. Thirty-five subjects were close-relaters, 16 non-close-relaters, and nine fell into neither category; there was a significant tendency for the non-close-relaters to be found in the goat group. The two groups designated as 'sheep' and 'goats' did of course differ enormously as regards their scores on the sheep-goat scale, and on every one of the ten questions comprising the scale. Less obviously, and unexpectedly, sheep and goats also differed on three of the six personality scales administered concurrently with the attitude-scale: t-tests for independent samples, with 58 df and two-tailed alpha of 5%, revealed that for Cattell's Factor H (Social boldness; mu = 5.5), the mean for sheep was 5.82, for goats

3.88, t= 3.71, 2p = 0.0003, omega-squared = 0.18*; for Social Skills (mu = 50), the sheep mean was 47.85, the goat mean 26.85, t= 3.36, 2p = 0.0014, omega-squared = 0.15; and for the MMPI Social Introversion Scale (mu = 50), the sheep averaged 53.38, goats 59.54, t= 2.21, 2p = 0.031, omega-squared = 0.06. Thus, it could be concluded that compared with sheep, the goats were more shy, more introverted, and more lacking in social skills, although the values of omega-squared indicate that the degree of association between personality and sheep-goat is not enormous. Nevertheless, these new findings do suggest rather more extensive differences between the two groups than merely attitudinal differences.**

Procedure.

(i) Re-test results. All subjects were initially administered the sheep-goat scale in March, 1977. The ESP-testing was carried out between July 4th and 21st of that same year, some 3-4 months later. It might thus be wondered to what extent the subjects' attitudes had changed over that time interval. It therefore seemed advisable to check that the subjects were still sheep or goats when it actually came time to test their psi-performance. (The

---

* Omega-squared may be thought of as a statistic representing the proportion of variance in the predicted variable which is accounted for by the predictor variable; that is to say, it describes the strength of association (or of the statistical relation) between the independent and dependent variables. For samples from two populations, each of which has the same true variance, a rough estimate of omega-squared is provided by

$$\text{est. } \omega^2 = \frac{t^2 - 1}{t^2 + N_1 + N_2 - 1}$$

If t is less than or equal to 1, then omega-squared is set to 0.0. Estimated omega-squared thus ranges from 0.0 to 1.0. (Hays,1963).

** The discovery of a small but consistent relationship between extraversion and the sheep-goat variable has been verified in several other sets of data, and a detailed report has been published in a psychological journal (Thalbourne & Haraldsson, 1980); see also Thalbourne (1981).

test-retest reliability of the sheep-goat scale would also be a useful piece of information to have in its own right.) For this reason, then, immediately prior to the telepathy test, each subject was re-administered the ten sheep-goat questions, together with an eleventh item asking them to estimate the likelihood of their showing ESP (see Appendix 3). The results of this re-test are given here briefly.

As in the original testing, the differences between subjects previously designated as 'sheep' or as 'goats', as regards their responses to the ten sheep-goat questions, were significant, generally enormously so; somewhat surprisingly, the only exception to this was the additional item regarding the estimated probability of displaying ESP: sheep were non-significantly more optimistic than were the goats, the modal response of the former being "50:50", whereas the majority of goats deemed the chances of psi occurring to be "unlikely". Test-retest correlations (Spearman) ranged from +.45 for the item on Paranormal Visions, up to +.88 for Personal Experience of ESP, and most were about +.70. Again, the (Pearson) test-retest correlation for the total scale-score was +.93 ($n = 60$, $2p = 3 \times 10^{-11}$). The group means for the scale were accordingly quite similar to those on original testing: 14.4 for sheep (compared with 14.8 before), and 4.4 for goats (a significant increase in positivity from their former mean of 2.8 ($t = 3.7$, $2p = 0.001$, 25 df), perhaps resulting from a dissonance-reducing mechanism.) It can thus be fairly safely concluded that the attitudes of the subjects proved reasonably stable over the 3-4 months between test and retest. In fact, only one subject -- originally a goat -- had altered her opinions so much that she was included in the sheep group for the actual ESP test.

(ii) Targets. Since all the subjects were new to the university, it seemed highly unlikely that they would have even heard about the Closeness Experiment conducted 12 months previously, much less have had contact with the subjects therein or learnt anything useful about any of the 120 targets

employed in that first study. It was therefore deemed satisfactory to use the 12 sets of targets previously prepared (see Appendix 1, sets 1 to 12), one set being used per (evening) session, with a maximum of three subject-pairs being tested (consecutively) in the same evening. Strict precautions were taken to prevent pairs tested later in the session from having any contact with pairs tested earlier in the session, lest information be passed on regarding the contents of the targets. In the next evening's session, a new set of targets would be used, and so on until all sessions were completed. Nevertheless, an unfortunate but unavoidable defect in experimental security was the fact that the experimenter had some knowledge as to what the targets were (by virtue of having set them up in the agent's room), and could thus have conceivably transmitted sensory cues to the subjects; acutely aware of this possibility, the experimenter was exceedingly careful not to say anything that would cue the participants in this way; indirect testimony to the effectiveness of this effort was the fact that, in the event, out of 300 attempts to reproduce the target-drawing, not one of the percipients ever achieved a perfect 'palpable hit'.

(iii) The ESP-test. Subject-pairs were individually tested on the ten-trial picture-guessing task without trial-by-trial feedback, as described in Chapter 2. The testing was carried out in two adjacent rooms, agent in one room (where also the experimenter sat quietly, reading or tabulating data and supervising operations where necessary), while the percipient sat at a desk in the other room, the door being closed and the fan set in motion (see diagram, p. 48). A noteworthy detail of procedure in this experiment was that the role of receiver was delegated to whichever of the two subjects was the more extreme in attitude: this was done because the correlation analysis had suggested that the attitude of the percipient was more determinative of ESP-score than was that of the agent, and thus the size of the sheep-goat/ ESP correlation should be enhanced by having maximally extreme sheep and

goats. Immediately following the tenth and final picture-guessing trial, but prior to receiving any feedback as to their performance, each subject completed the standard post-experimental questionnaire (Appendix 7), in order to record how they had felt and how well they thought they had done, etc.

(iv) Judging. A total of 44 judges were employed to rank-evaluate the response-material, in the manner described in Chapter 2. Most of the judges were unpaid volunteers from the sample of 235 Psychology students referred to above. The 30 sets of response-drawings (that is, ten responses from each of the 30 percipients) were preferentially ranked, three times and independently, against each of the ten respective targets used. For the resulting total of 300 'triple-evaluations', the concordance between the three rank-orderings was significantly high (that is, W* equalled or exceeded 0.627) on 55% of occasions, and thus was non-significant (fell short of 0.627) on 45% of occasions; grand mean Kendall's W coefficient of concordance was 0.64, and the consistency of this figure is demonstrated by the fact that the mean W for sheep was 0.65, and for goats 0.62 -- a non-significant difference.

RESULTS

I. The Post-Experimental Questionnaire.

(i) For the two multiple-choice questions (regarding the subject's belief as to whether or not ESP had occurred, and as to whether there were any psi-inhibitory factors), frequency counts were made of the number of subjects falling into each of the three response-categories (namely "yes", "uncertain", "no"). These frequencies were broken down by the sheep-goat dichotomy, and also by the agent-percipient dichotomy, the resulting 3 x 2 contingency tables being subjected to the Chi-Square test, with 2 df. Sheep did not differ significantly from goats regarding their belief as to whether ESP had occurred, but there was a significant agent-percipient difference ($\chi^2 = 8.19$, $p = 0.02$): percipients were more inclined to say "no" (47%) than were agents

---

* See footnote, Chapter 6, p.

(13%), and agents were more inclined to say "uncertain" (80%) than were percipients (47%); comparatively few subjects (6% of percipients, 7% of agents) thought that ESP had occurred. By way of contrast, agents and percipients did not differ in their responses to the question of whether there might have been a psi-inhibitory factor, but there was a significant association with the sheep-goat variable ($\chi^2 = 7.86$, $p = 0.02$): sheep were more likely to say that there was such a factor (74%) than goats (39%), with only 15% sheep saying "no" as opposed to 42% goats. The two most commonly expressed opinions to account for possible lack of success were (i) lack of familiarity with their experimental partner, and (ii) the bright lighting of the laboratory rooms, followed closely by such factors as "the distracting effect of the occasional background noises", "fatigue", and "difficulty in concentration". By way of summary, we may say that the percipients were less confident than were agents that ESP had occurred (though, surprisingly, sheep were no more confident than were goats); and that sheep were more inclined to mention conditions which they deemed psi-inhibitory, whereas goats had less motive to do this, probably because they lacked belief in or experience of psi phenomena in any case.

(ii) Five rating-scales were administered, the purpose being a) to test the efficacy of the experimental instructions relating to the mental state required during the ESP test, and b) to ascertain whether sheep and goats, and senders and receivers, were comparable on mood-state variables. The check-marks made on these scales (see Appendix 7) by the subjects were transformed into scores on a 1-6 scale, and t-tests for independent samples, and one-way analyses of variance (for randomized groups) were used for two-group and four-group comparisons respectively. In general, the results turned out as had been hoped: no significant differences between sheep and goats, or between senders and receivers, and overall mean scores at the expected end of the scale: on average, the 60 subjects rated themselves as

having been quite relaxed (mean = 4.3), very interested (5.0), fairly calm (2.5 on the Agitation scale), quite enthusiastic (4.8), though somewhat uncertain (4.0), as opposed to being 'tense', 'bored', 'agitated', 'unenthusiastic' and 'confident'. However, there was a significant interaction effect between belief and role on the Enthusiasm scale ($F = 4.09$, 3 and 56 df, $p = 0.01$), with goat percipients being the least enthusiastic, sheep percipients the most enthusiastic, and the agents (sheep and goats) being almost identical to each other but half-way between the other two groups. This difference is to be expected, by virtue of the fact that enthusiasm is correlated with the level of belief in psi ($r = +.28$, $2p = 0.031$): it was the goat percipients who were most skeptical (mean ten-item sheep-goat scale-score being 3.0), and the sheep percipients who were the most believing (mean = 15.2).

II. The ESP Test.

(i) The Sheep-Goat Effect. The major hypothesis of interest was that there would be a significant and positive Pearson correlation between, on the one hand, the two ESP-scores mean rank and hits, and on the other hand, the sheep-goat scale-scores of the subjects, particularly the percipients. The hypothesis was not, however, supported, and in fact even the direction of the correlation was negative, contrary to prediction. For the 30 percipients, the correlation with total number of hits — previously +.52 — fell to -.04, and with mean rank fell from -.51 to +.09. For the 30 agents, the correlation with hits was -.14 (previously +.33) and with mean rank was +.20 (previously -.20). Thus, the correlations obtained in this experiment were quite non-significant, and the null hypothesis was retained of no relationship whatsoever between attitude and ESP-score. To state the obvious: the effect previously observed, failed to replicate. The failure does not seem explicable in terms of a marked difference in score-range between the original and the follow-up study: here, total number of hits ranged from 1-8 (previously 2-8), mean rank 3.50-7.37 (previously 4.23-7.30), and the 11-item sheep-goat scale from 1-22 for percipients (1-21 before), and 5-20 for agents (3-24 previously).

It was to have been expected that if the sheep-goat correlation had held, then a statistical consequence of this would probably have been that the grand mean ESP-scores for sheep would have been significantly higher than those for the goats. But, consistent with the low non-significant negative correlations actually obtained, the grand mean scores were almost identical, with the goats non-significantly higher than the sheep: the grand mean rank for sheep was 5.49, for goats 5.41, t = 0.25, 28 df, 2p = 0.801; these means each deviate from chance (5.50) to a non-significant positive degree; again, the mean number of hits obtained by sheep was 4.76, by goats 4.92, t = 0.25, 28 df, 2p = 0.801, these means deviating from MCE of 5 in a non-significant negative direction. The mean absolute-deviation measure, about which no specific expectations had been entertained, came the closest to yielding a significant difference: whereas MCE is approximately 1.90, the grand mean for sheep was above chance at 2.02, for goats below at 1.77, t = 1.78, 28 df, 2p = 0.091.

(ii) Individual Scoring. The average rank-scores obtained by each subject-pair on each of the ten trials in the ESP-test, may be found in Appendices 12 and 13. Mean scores and other summary statistics for individual pairs are to be found in Tables 3.3 and 3.4. In each table, the mean ranks are arranged in ascending magnitude. For mean ranks, the hypothesis is two-tailed, while the probabilities yielded by the Randomization Test are inherently one-tailed; hence, alpha being 0.05, the probability of a pair's mean rank must be 0.025 or less in order to be considered significant. The hypothesis for mean absolute-deviations, however, is one-tailed (that is, only significantly large variance is being examined); alpha again being 0.05, the probability of a pair's mean absolute-deviation must accordingly be less than or equal to 0.05 to be significant. For the hits-variable, the hypothesis is two-tailed, and in order to achieve a significant score, a pair must obtain fewer than 2 or more than 8 hits.

Explanatory Notes: Mean Ranks are found by averaging the average rank-scores of a pair's ten response-drawings; the range of average rank-scores is 1.00 (most similar) to 10.00 (most dissimilar), with a mean chance expectation of 5.50. The '1p(Mean)' value represents the chance-probability as estimated by the Randomization Test; for mean ranks less than 5.50, p is the probability of obtaining a mean smaller than or equal to the actual obtained mean; for mean ranks greater than 5.50, p is the probability of finding a mean greater than or equal to the obtained mean.

A response-drawing is scored as a 'hit' if its average rank-score is less than 5.50, or as a 'miss' if the average rank-score is greater than 5.50. The total Number of Hits is therefore obtained by simply summing the number of hits obtained over the ten telepathy-test trials.

Mean Absolute-Deviations represent the average amount by which the average rank-scores of a pair's ten response-drawings deviate (in either direction) from the expected mean of 5.50; the range is 0.17 to 4.50; here, '1p(Mean)' represents the Randomization Test probability of a mean absolute-deviation occurring which is as great as or greater than the obtained mean.

Table 3.3. Mean Ranks, Hit-Scores, and Mean Absolute-Deviations for the Sheep.

	Average-Rank			Number	Absolute-Deviation		
Pair	Mean	S.D.	1p(Mean)	of hits	Mean	S.D.	1p(Mean)
50	4.17	2.37	0.052	7	2.37	1.34	0.347
54	4.60	1.74	0.111	7	1.67	1.02	0.786
55	4.73	1.55	0.130	5	1.33	1.11	0.903
41	4.80	1.81	0.148	7	1.60	1.10	0.669
33	4.97	2.13	0.238	6	1.80	1.25	0.623
60	4.97	2.53	0.228	6	2.33	1.11	0.108
32	5.33	2.69	0.413	5	2.33	1.34	0.207
35	5.43	2.77	0.481	4	2.47	1.26	0.113
38	5.53	2.22	0.470	4	1.83	1.25	0.667
43	5.57	2.51	0.454	4	2.13	1.32	0.234
52	5.73	2.26	0.382	6	1.97	1.15	0.499
46	5.83	2.05	0.352	4	1.87	0.90	0.634
39	5.93	1.98	0.290	4	1.77	1.00	0.682
40	5.97	2.87	0.263	4	2.60	1.31	0.027
45	6.07	2.50	0.215	5	2.27	1.20	0.252
36	6.33	1.98	0.142	1	1.73	1.27	0.794
57	7.37	1.70	0.005	2	2.27	1.12	0.257
Grand Total	5.490	0.76	-	4.765 (s.d.=1.68)	2.020	0.36	-

Table 3.4. Mean Ranks, Hit-Scores, and Mean Absolute-Deviations for the Goats (n=13)

Pair	Average-Rank			Number of hits	Absolute-Deviation		
	Mean	S.D.	1p(Mean)		Mean	S.D.	1p(Mean)
34	3.50	1.70	0.008	8	2.27	1.33	0.362
56	3.67	1.96	0.004	8	2.30	1.38	0.070
42	4.87	1.86	0.196	7	1.67	1.04	0.807
58	5.10	1.76	0.268	6	1.40	1.14	0.869
51	5.47	1.56	0.506	5	1.20	0.99	0.988
53	5.70	2.42	0.399	4	2.00	1.38	0.227
61	5.80	1.56	0.333	4	1.37	0.82	0.854
48	5.87	1.96	0.275	5	1.67	1.10	0.535
47	5.90	1.61	0.277	4	1.47	0.78	0.842
44	6.03	2.34	0.226	4	2.07	1.22	0.309
37	6.10	1.28	0.201	3	1.17	0.80	0.981
49	6.13	2.50	0.138	3	2.13	1.46	0.374
59	6.23	2.56	0.150	3	2.30	1.34	0.202
Grand Total	5.413	0.91	-	4.923 (s.d.=1.80)	1.769	0.43	-

From Tables 3.3 and 3.4 it can be seen that despite the lack of overall group significance, five of the 30 subject-pairs achieved significance on one or other of the three measures of ESP-performance. Two pairs of goats scored significantly positively: one pair, 34, obtained a mean rank of 3.50 (Randomization-Test estimate of probability being 0.008), and pair 56, averaged 3.67 (1p = 0.004). Two pairs of sheep scored significantly negatively: pair 57 had a mean rank of 7.37 (1p = 0.005), while pair 36 (who, incidentally, are the only twins ever formally tested by this experimenter), scored one hit and nine misses — an event which has a one-tailed binomial probability of 0.011. Finally, one pair of sheep (pair 40) had a mean absolute-deviation of 2.60, and the one-tailed probability of this value exceeding 1.90 is 0.027. Though it tends to rub salt into the wound that individual pairs of sheep or goats should score significantly opposite to the direction predicted for the groups as a whole, nevertheless it is some balm to know that in the case of the mean rank measure, the probability that three of more pairs will score significantly at the (average, two-tailed) level or 0.0113, is itself significant, the p being 0.0047.

(iii) Serial-Position Effect. One of the predictions that had been advanced in the earlier Closeness Experiment (Thalbourne, 1976) was that the level of scoring displayed by the two subject-groups would change significantly over time, that is, over the ten trials in the experiment. This prediction was in fact confirmed: repeated measures analysis of variance, performed on the combined group scores, revealed a significant effect of trials, for both the mean rank measure ($F = 1.984$, 9 and 270 df, $p = 0.041$) and for the hits-measure ($F = 2.339$, 9 and 270 df, $p = 0.015$). Trend analysis, using the method of orthogonal polynomial coefficients, suggested that the quadratic component was significant for the hits ($F = 4.121$, 1 and 270 df, $p = 0.043$), that is, that the ten trial-means of the combined groups were best described as following a U-shaped trend. A similar analysis was carried out

FIGURE 3.1 : TRIAL-BY-TRIAL SCORING FOR THE FIRST SHEEP-GOAT EXPERIMENT

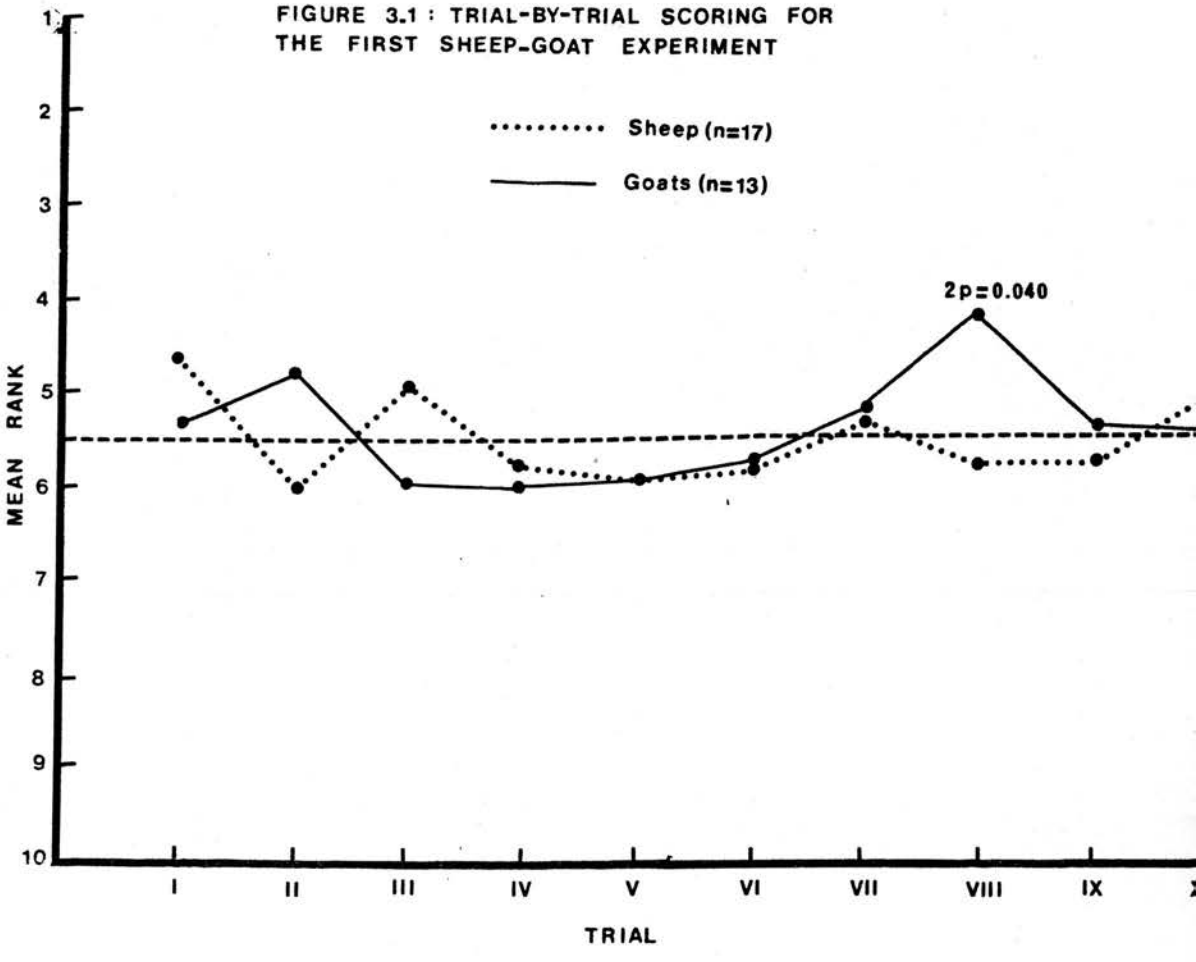
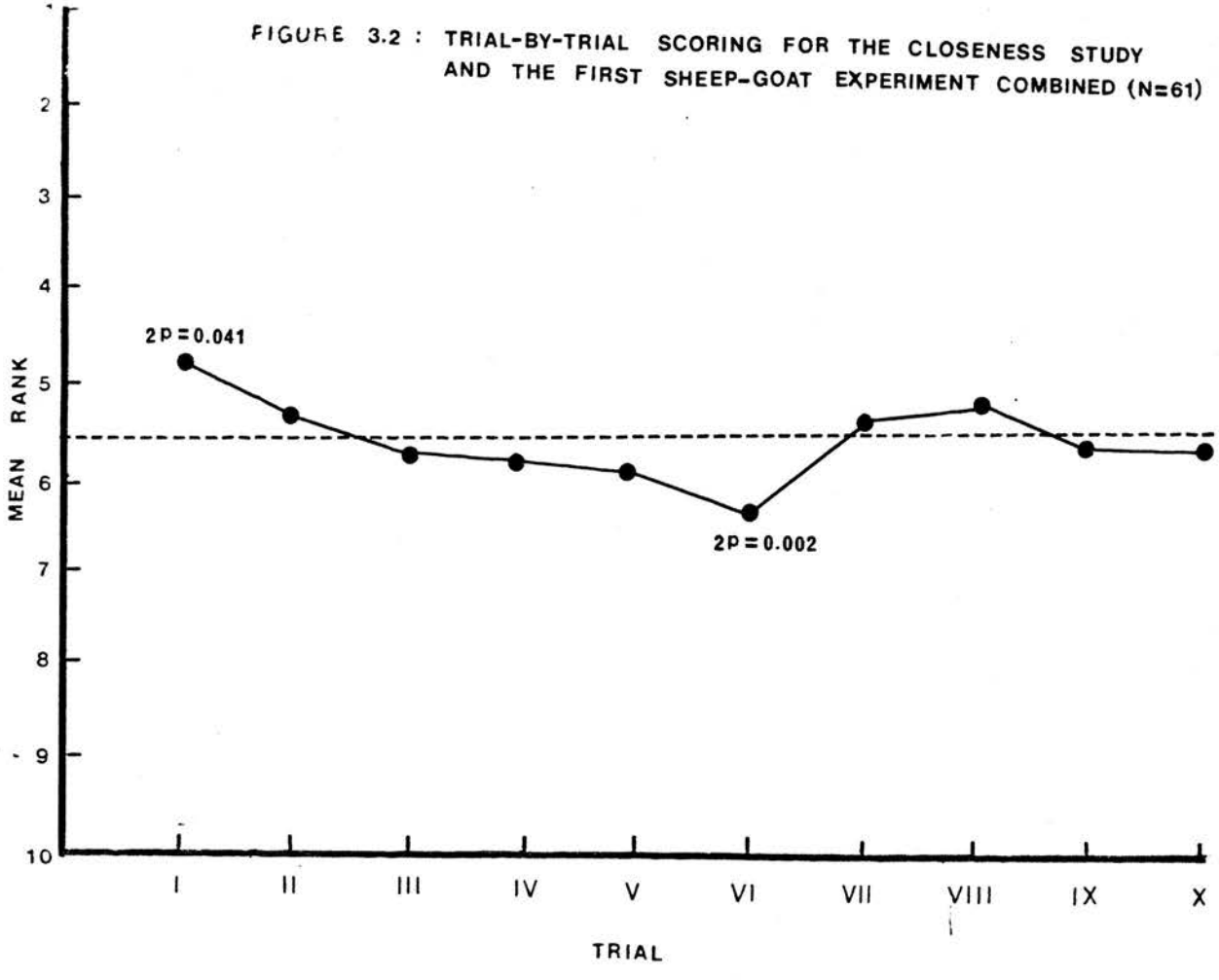


FIGURE 3.2 : TRIAL-BY-TRIAL SCORING FOR THE CLOSENESS STUDY AND THE FIRST SHEEP-GOAT EXPERIMENT COMBINED (N=61)



for the present Sheep-Goat Experiment, in the hope of finding this 'serial-position effect' a second time. Figure 3.1 displays in graphic form the mean rank-scores of sheep and goats for each of the ten trials. Though the graph vaguely resembles a U-shape, statistical analysis failed to reveal any significant evidence for a change in level of scoring over time, or for any conformance to a mathematical trend. However, when the scores of the 30 sheep/goat pairs were combined, in post hoc fashion, with the scores of the 31 close/non-close pairs, repeated measures analysis of variance for the combined N of 61 gave a just-significant effect of trials ($F = 1.894$, 9 and 540 df, $p = 0.050$), and trend analysis also yielded a significant quadratic component (for mean rank, $F = 5.744$, 1 and 540 df, $p = 0.017$; for hits, $F = 4.313$, 1 and 540 df, $p = 0.038$). Figure 3.2 represents the combined mean rank-scores for all 61 subject-pairs, for each of the ten trials. There thus seems to be some indication that there is a slight but consistent serial-position effect occurring, that is, an effect in which the level of scoring on a particular trial is (weakly) related to the position which that trial occupies in the sequence of ten. It was something to watch for in future experiments.

Conclusion. Despite the possible outstanding psi-performance of a few isolated subject-pairs (ironically, the sheep psi-missing and the goats psi-hitting), together with a very faint suggestion of a serial-position effect, it must be conceded that this study was an abject failure to elicit psi in the predicted fashion: knowlege of a person's score on the Sheep-Goat scale was not sufficient to enable us to predict their score on the ESP test. There thus seemed, on a straightforward level, to be no evidence whatsoever for the Sheep-Goat Effect. In Chapter 4, we explore possible reasons for this null result.

CHAPTER 4

The Second Sheep-Goat Experiment

CHAPTER 4

"Faith is the substance of things hoped for,
the evidence of things not seen"

The Epistle of St. Paul the Apostle
to the Hebrews, XI, 1

PART I

Given the rather miserable null results of the sheep-goat experiment reported in the previous chapter, the suggestion was made to the writer that, most likely, the original significant sheep-goat correlations were spurious, that is, artifacts produced as a result of over-analyzing the data. And it was said that given the unknown amount of interrelatedness between the variables, there was no a priori way of telling whether more correlations had been significant than would be expected by chance. Though this was an understandable reaction, the author nevertheless felt convinced that the number and magnitude of the original significant correlations was just too great for it to be an artifact. This first part of the chapter will be devoted to demonstrating how some persuasive evidence was obtained for the truth of this conviction. However, let it be said straightaway that the author did not totally vindicate himself: in the course of demonstrating that there was something more than an artifact, there also came to light a ratiocinative error that had led the experimenter astray; this knowledge clearly illuminated why the sheep-goat experiment had been a failure, and showed the way to a better experiment, which forms the substance of Part II.

In essence, what was done was a series of control tests — three in number — and performed by computer. These will be termed 'the Shuffle Controls'. It was reasoned that if the original significant correlations were genuine, then, logically, what this meant was that the 'psychological structure' of any given subject (particularly their belief in, and supposed experience of, the paranormal) was systematically related to the level of their ESP-score.

At the same time, the psychological structure of that subject should not have any influence on the ESP-score of somebody tested at a different time in the experiment. So what was done was to re-allocate the ESP-scores — 'shuffle them about', as it were — using some random process to decide to which subject-pair each ESP-score should be re-allocated.* Any correlation that should then exist between psychological variables and ESP-scores, ought then to be due to chance factors: it would seem a very strange thing indeed if the psychological structure of percipient **x**, tested on Wednesday, had any influence on the ESP-score of percipient **y**, tested on the previous Monday!

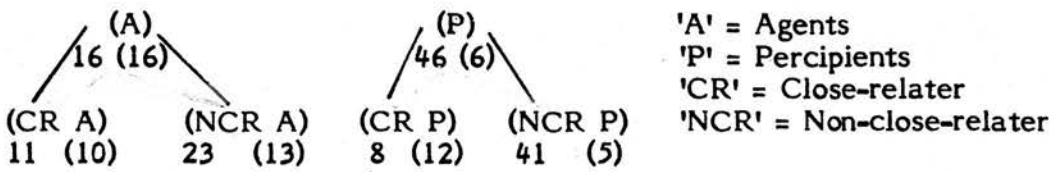
The beauty of this approach was that the degree of relatedness between the predictor variables remained the same (since they were being transported en bloc across subjects, as a complete set), and therefore any results of a Shuffle test were strictly comparable to those of the correlation analysis originally carried out on the 'unshuffled' data. In a way, the process resembled a Randomization Test; it was a method of ascertaining how many correlations one would expect to be significant when only chance (presumably) was operating. A total of three such Shuffle Controls were carried out, in order to give Tyche free rein to do her worst: if, on none of the three occasions, the control tests produced as many significant correlations as did the actual test, then it could more plausibly be argued that the latter contained evidence for a real effect.

Now recall that the actual test yielded a total of 145 significant correlations, or just over 12%. The maximum number of significant correlations ever obtained on any of the three control tests, was 45, or 3%. But even more interesting is the distribution of these chance-produced correlations between the six subject-groupings. In the schematic representation below, the figures in brackets represent the highest number of significant correlations

---

* Shuffled counters were deemed to be adequate for this purpose.

ever obtained on any of the three control tests for that particular grouping of subjects; these values thus constitute the strictest and most conservative 'test of significance' possible under the circumstances of only three control tests. The unbracketed figures refer to the number of significant correlations obtained under the original analysis, for purposes of comparison (cf. p 97, Chapter 3).



On the assumption that the bracketed figures give a reasonable estimate of what may be expected by chance, it can immediately be seen that the figure for all agents on the actual correlation analysis, namely 16, is precisely what would have been predicted on the null hypothesis; it is a similar story for that subgroup of agents who were close-relaters (11 actual versus 10 control correlations); for the non-close-relater agents, however, the actual number of significant correlations is approaching double that expected by chance, and this rather suggests a real effect. As regards percipients, the situation is very clear-cut indeed: as a group, the 31 receivers show a very large excess of observed over expected significances (46 actual as opposed to 6 control); this excess appears to be entirely attributable to the non-close-relater percipients. Hence, the Shuffle Controls provide very strongly suggestive evidence that something extrachance was occurring in the original experiment: however, that 'something' was confined to the non-close-relater pairs, and, within them, most certainly the percipients but perhaps also the agents as well.

It was subsequently suggested to the author that a more elegant way of approaching the problem would be to make use of the multivariate tool known as 'multiple regression': such a method would yield an index -- the multiple correlation coefficient -- which measures the overall success with which the

psychological variables, as a set, predicted ESP-score. For various technical reasons, however, it was not possible to use so large a set as 33 variables simultaneously to predict the direction of psi-scoring.* It was therefore necessary to reduce the larger number of variables to a smaller set, and this was achieved by subjecting the 33 psychological scores of the 62 subjects to principal-components factor analysis, with Quartimax rotation. Principal-components factor analysis is a relatively straightforward method of transforming a given set of variables into a new set of composite variables ('factors' or 'principal components') that are orthogonal to (that is, uncorrelated with) each other: the factors are defined as exact mathematical transformations of the original data. Quartimax rotation is a method whereby the 'complexity' of each variable is minimized by making clear which of the original variables are important for the factor: it rotates the initial factors in such a way that a variable loads high on one factor but almost zero on all others. The computer-program used (Kim, 1975) retained only components with eigenvalues greater than or equal to 1.0 — a criterion which ensures that only components accounting for at least the amount of the total variance of a single variable will be treated as significant.

It was found by experiment that very similar factors emerged whether the entire set of 33 psychological variables was factor-analyzed in one go, or

---

* (i) Firstly, many of the predictor variables are only on an ordinal scale, and multiple regression requires them to be on interval or ratio scales. (2) Secondly, and more seriously, the statistical significance of the multiple correlation coefficient is tested by means of an F-test, with k and (N - k - 1) degrees of freedom, where k is the number of predictor variables and N the sample-size. Thus, if the second degrees of freedom is not to be zero or negative, then the sample-size (N) must be greater than (k-2); this condition is clearly not met for N=13 (in the case of the non-close-relater agents or percipients) if k, the number of predictors, is 33. (3) And third is the problem of 'multicollinearity', which refers to the situation where some or all of the predictor variables are very highly interrelated (which is true of many of our 33 psychological variables). When multicollinearity is extreme, there is, unfortunately, no acceptable way to perform regression analysis using the given set of independent variables. Kim & Kohout (1975) suggest that a way to cope with this problem is to create a new variable which is a composite scale of the set of highly intercorrelated variables and use the new scale variable in the regression equation in place of its components.

the 33 were subdivided into their 'logical' groupings (namely, Social Intelligence, Personality, Mood, and Sheep-Goat variables) and these four subgroups factor-analyzed separately. Thus, the latter method was preferred, for reasons of convenience. The process yielded a total of nine factors, which are listed down the left-hand side of Table 4.1 (together with the correlations between them and the ESP-score Hits, for each of the six subject-groupings.)

Table 4.1

Pearson Correlations Between Factor-Scores and Number of Hits out of Ten

	Agents	CR A	NCR A	Percipients	CR P	NCR P
Social Intelligence	-.35	-.45	-.29	-.21	-.18	-.12
Extraversion	+.01	-.02	-.08	+.44**	+.46	+.25
Anxiety	+.25	+.22	+.31	-.12	+.03	-.32
Enthusiasm	+.09	+.13	-.25	+.03	+.04	-.16
Psychic Experiences	+.17	+.01	+.29	+.46**	+.22	+.73**
Other-worldliness	+.47**	+.49*	+.43	+.07	-.12	+.67*
Experience of Telepathy	+.15	-.10	+.31	+.36*	+.44	-.03
Sheep-Goat (Experimental)	-.02	-.18	+.21	+.05	-.20	+.34
Sheep-Goat (Theoretical)	-.16	-.08	-.21	+.33	+.29	+.30

* : $2p \leqslant 0.05$ ** : $2p \leqslant 0.01$

'CR A' = 'Close-Relater Agents'
'CR P' = 'Close-Relater Percipients'
'NCR A' = 'Non-Close-Relater Agents'
'NCR P' = 'Non-Close-Relater Percipients'.

Comparison of the names of the factors here, and the original variables in Table 3.1 (Chapter 3, p. 88), will show that: the four original social intelligence variables reduced to a single factor (accounting for 66% of the total variance in those four scores); similarly, the six personality scales were not each measuring different things, but rather just one dimension which has been named 'Extraversion'

(58% of the variance); the five mood-measures reduced to two factors (totalling 73% of the variance); and the attitude-scales and items together yielded five distinct underlying factors (accounting for 75% of the total sheep-goat variance). Although these nine factors are not perfectly orthogonal to each other, they nevertheless intercorrelate to a much lesser degree than do some of the original 33 variables, thus alleviating the problem of multicollinearity. Table 4.1 thus represents a condensation, or distillation, of the bulk of the information in the previous Table 3.1.

Having done this, it was then possible to perform a multiple regression for each of the six subject-groups, using the nine factors to predict ESP-score. This technique would yield a multiple correlation coefficient, representing the strength of the linear relationship between the set of nine predictor variables, and ESP-score. Eighteen regressions were carried out (three ESP-scores x six subject-groupings), of which only one yielded a significant multiple correlation: this was for the non-close-relater percipients, when mean rank was the predicted ESP-score: multiple $r = .987$, $F = 12.38$, 9 and 3 df, $p = 0.031$ (or 1 in 32). This correlation sounds very high, but such large values are necessary when the sample-size is as small as 13. Lest it be thought that this sole significance is spurious (only one in 18 analyses being significant), let it be said that in a manner analogous to the Shuffle Controls, three sets of control multiple regressions were performed (essentially, using one subject's factor-scores to predict the ESP-scores obtained by <u>another</u> subject); yet out of a total of 54 analyses (that is, 18 x 3), <u>none</u> was significant.

Multiple regression also indicated <u>which</u> of the non-close-relaters' factor-scores contributed most to the significance of the overall multiple correlation. More technically, it indicated which specific beta weights (i.e. standardized regression coefficients) were non-zero. The significance of the beta weights can be tested using a further F-test, with 1 and (N-k-1) degrees of freedom,

Table 4.2 Correlations Between Mean Rank and the Sheep-Goat Scale and Component Items, Closeness Study

	AGENTS			PERCIPIENTS		
	All A (n=31)	CR A (n=18)	NCR A (n=13)	All P (n=31)	CR P (n=18)	NCR P (n=13)
Sheep-Goat Scale 3	-.20	+.06	-.68**	-.51**	-.34	-.79***
Belief in ESP	-.03	+.06	-.17	-.42*	-.12	-.80***
Personal Experience of ESP	-.01	-.02	-.12	-.42*	-.43	-.54
Believe Self to be Psychic	+.06	+.31	-.40	-.53**	-.40	-.73**
Have had a Hunch	-.18	+.04	-.51	-.22	-.12	-.30
" " " Premonition	-.07	+.22	-.50	-.24	-.22	-.33
" " " Precognitive Dream	-.40*	-.41	-.49	-.44*	-.22	-.70**
" " " Vision	+.03	-.10	+.12	+.03	+.05	—
Believe in Life After Death	-.20	+.19	-.44	-.06	-.01	+.03
" " Spirit Contact	-.20	-.13	-.19	-.34	-.10	-.52
Telepathic Experience with Partner	-.07	-.07	—	-.16	-.29	—
Telepathic Experience with non-Partner	-.13	+.06	-.58*	-.58 †	-.69***	-.40
Have Had a Telepathic Experience	-.18	0.00	-.58*	-.51**	-.60**	-.40
Estimated Probability of Success	-.27	-.21	-.32	-.34	-.06	-.63*
Belief that ESP Occurred	+.06	+.27	-.41	-.11	-.23	-.04
" " was inhibiting factor	+.17	+.22	+.10	-.13	-.21	-.07

Note:
* : $2p \leq 0.05$
**: $2p \leq 0.01$
***: $2p \leq 0.001$
† : $2p = 0.00068$

'A' = 'Agents'
'P' = 'Percipients'
'CR' = 'Close-Relater'
'NCR'= 'Non-Close-Relater'

in the present case 1 and 3 df: in order of significance, these were: Other-worldliness (that is, belief in survival of bodily death and in the possibility of contact with spirits of the deceased: $F = 25.86$, $p = 0.015$); sheer number of reported spontaneous Psychic Experiences ($F = 20.08$, $p = 0.021$); Social Intelligence ($F = 18.76$, $p = 0.023$) and Sheep-Goat attitude in relation to the subject's performance in the ESP experiment ($F = 17.95$, $p = 0.024$).

There thus seemed to be an additional piece of evidence to suggest that there really was a sheep-goat effect in the original data of the Closeness Study, although the multiple-regression analysis implied that the effect was confined to non-close-relater percipients (and indeed only their scores on the mean rank variable).

Thus, as a result of the three Shuffle Control tests, and the multiple-regression analyses, two conclusions were drawn: (1) it had been correct to suppose that there was a non-chance effect in the original correlations, but (2) it was incorrect to have supposed that the effect would occur in a basically unselected group of percipients: the percipients had to be non-close-relaters, or the effect would not occur. So instead of a straightforward sheep-goat effect, what had been found was an 'interactive sheep-goat effect', that is, an effect whose occurrence depends critically upon an additional variable, namely, whether or not the subject is a non-close-relater.

When the author returned to the original correlations, it then became obvious how an error in reasoning had been made. (For the convenience of the reader, see Table 4.2, which displays the original correlations between the attitude variables and the 'mean rank' ESP-measure; bear in mind that high values of mean rank indicate negative scores, whereas low values represent positive scores; thus, the sign of the correlation is opposite to that for the 'hits' variable (cf. Table 3.1.)) As can be seen from Table 4.2, the significant correlations tend to be loaded on the non-close-relaters: the correlations for the close-relaters are, almost completely, non-significant

(which is to say, essentially zero in the population, though in samples unstable within certain limits of fluctuation.) Looking particularly at the results for the percipients (columns 4 through 6), the correlations for the undifferentiated group of 31 receivers (column 4) constitute what we may term a 'dilution effect': the significance and size of the correlations for the non-close-relater percipients has been reduced — 'watered down', or diluted — by the lower, non-significant correlations of the close-relater percipients; to have combined the two subgroups, indiscriminately, is only to have added so much irrelevant 'noise' to the real effect. And the really crucial point is that given that the non-close-relater correlations were significant, the significance of the combined group depends very much upon the magnitude of the close-relater correlations, even though the latter are non-significant. For even given that the close-relater correlation is essentially zero, it can still, due to sampling error, fluctuate either side of zero, positively or negatively, within the limits set by chance. If the close-relater correlation happens by chance to turn out moderately high in the direction opposite to that of the non-close-relaters, then the effect of combining the two correlations is similar to two opposing forces meeting and cancelling each other out, as illustrated below in Table 4.3

Table 4.3

Hypothetical representation of the statistical 'Dilution effect'.

NCR (n=13)	CR (n=18)	Combined (n=31)
r = +.80, p=sig.	+.24, p=n.s.	+.52, sig.
+.80 "	0.00, n.s.	+.40, n.s.
+.80 "	-.24, n.s.	+.10, n.s.

Thus, in the original analysis, the fact that the Sheep-Goat Scale/Hits correlation for all percipients (namely, r = +.52, 2p= 0.003) was still

significant after the data from the non-close-relaters (r = +.80, 2p = 0.001) had been pooled with that from the close-relaters (r = +.24, 2p = 0.344) was merely a lucky fluke. The correlation for the close-relaters could as well have been slightly negative, and have thus caused the overall correlation for percipients to have been non-significant (as suggested in the third row of Table 4.3 above). The author had failed to take into account the warning that the presence of a close relationship had a diluting effect on the correlation and should have entirely excluded such dyads from the follow-up sheep-goat study.

There is even a little evidence to suggest that such a dilution effect also occurred in the replication experiment (which might otherwise have seemed a predictive failure). Out of a total of 60 subjects in that experiment, there happened to be 16 subjects who were non-close-relaters (or at least who had reported themselves to be such, some three months prior to the ESP-test): 12 of these people had been delegated the role of agent, four of percipient; (of the 30 pairs of subjects, there were three pairs of non-close-relaters, while the remaining ten non-close-relaters were paired with a person who was not a non-close-relater). Since there were only four non-close-relater percipients, this was rather too small a sample to expect the sheep-goat correlation to be significant. Nevertheless, the correlations were most definitely in the right direction: r = +.43 for the hits variable, and -.35 for the mean rank. (In fact, the subjects' predictions as to how well they thought they would do on the ESP task correlated r_s = -.63 for mean rank (2p = 0.37) and +1.00 (2p = 0.001) for the hits variable!) The Sheep-Goat Scale correlations for the 12 non-close-relater agents were even more encouraging, being significant for mean rank (r = -.58, 2p = 0.047). Though lesser in magnitude than the correlations obtained in the first study, these smaller values, when combined across the two samples, nevertheless contribute to a highly significant set of correlations, as can be seen from Table 4.4.

Table 4.4

Pearson correlations between ESP-score and the Sheep-Goat Scale, for non-close-relater subjects only, first and second studies compared and combined

		MEAN RANK		HITS	
	STUDY	r	2p	r	2p
agents	1st (n=13)	-.68	0.010	+.57	0.040
	2nd (n=12)	-.58	0.047	+.34	0.282
	Combined(n=25)	-.64	0.001	+.47	0.018
percipients	1st (n=13)	-.79	0.001	+.80	0.001
	2nd (n=4)	-.35	0.648	+.43	0.574
	Combined(n=17)	-.76	0.0004	+.70	0.002

For various reasons, one would not want to place undue emphasis on these post hoc results: (1) multiple regression failed to suggest the occurence of any extra-chance effect among the non-close-relater agents of the first study (though the Shuffle Controls did suggest otherwise); (2) the analysis assumes that the variable conducing to the sheep-goat effect is 'being a non-close-relater', rather than (say) 'not having a close relationship with one's partner'*; and (3) even if 2) were correct, we cannot be absolutely certain that the subjects were still non-close-relaters by the time of the ESP-test (although from other research it appears that this would be highly likely).

---

* If, for the sake of the present post hoc analysis, we assume instead that the crucial variable is 'not having a close relationship with one's partner in the experiment', then it is possible to abstract from the 30 pairs comprising the Sheep-Goat Study, 13 pairs in which it can be reasonably assumed that this was the state of the agent-percipient relationship, by virtue of the simple fact that at least one of the persons in the pair was a non-close-relater: in three pairs both agent and percipient were non-close-relaters; in nine other pairs, the agent alone was a non-close-relater; and in one remaining pair, only the percipient was a non-close-relater. We thus have a total of 13 agents who were not in a mutually close relationship with their partner, and (it logically follows) 13 non-close percipients. For the group of 13 non-close agents, the sheep-goat/ESP correlation is -.54 (2p = 0.057) for mean rank, and +.29 for hits (2p = 0.328); for the 13 non-close percipients, the values are -.69 (2p = 0.009) and +.45 (2p = 0.121), respectively. These results are thus as encouraging as are those for non-close-relaters only, and we shall have more to say about this alternative hypothesis in Part II below.

Nevertheless, despite these caveats, Table 4.4, in conjunction with the clearer thinking resulting from re-examination of the original correlations, did seem to permit cautious optimism. But of course the real proof of the pudding was to conduct a second sheep-goat experiment, using a new group of non-close-relaters. This brings us to Part II of this chapter: the formal attempt to replicate this apparent 'interactive sheep-goat effect'.

PART II

Prologue. If we assume that pairs of non-close-relaters will exhibit a significant correlation between attitude and ESP-score, then a rather subtle question comes to mind regarding the causation of this effect. Does the ESP effect occur because the subjects are that unusual sort of person we call the 'non-close-relater', or is it due simply to the fact that the pair fail to share an extremely close relationship with each other ? The author's own psychological research (unpublished, but alluded to in the previous chapter, p. 87) has shown that non-close-relaters tend to differ somewhat from the population at large as regards their personality-structure, in that they tend to be more shy, more introverted and more lacking in social skills. So the cause of the observed sheep-goat effect may be some personality characteristic of non-close-relaters as such, or it may simply be the social-psychological situation between the agent and percipient, that is, the fact that they do not know each other very well, and that there does not exist between them a close emotional bond.

We therefore have two possible candidates for the role of relevant causal antecedent, one of which may be called the 'personality hypothesis', and the other, the 'relationship hypothesis'. Different empirical consequences are implied by each hypothesis when it comes to predicting ESP between persons who are strangers to each other but not non-close-relaters. If the relationship hypothesis is correct, that is, if the absence of a close relationship is sufficient to produce a significant

sheep-goat effect, then we would expect that (i) firstly, pairs of non-close-relaters would exhibit the effect; (ii) secondly, that pairs of mere strangers would also exhibit the effect; and (iii) thirdly, there would be no difference in the magnitude of the effect shown by each of the two groups. But, on the other hand, if it is the personality hypothesis which is nearer the truth, that is, if the presence of some attribute of non-close-relaters quite apart from their not having a close relationship with their partner, is sufficient to produce the correlation, then we would expect that (i) pairs of non-close-relaters would show a significant correlation; (ii) pairs of strangers would not produce that correlation; and (iii) there would probably be a significant difference between the size of the two correlation coefficients, that of the non-close-relaters being larger than that of the strangers.

Thus, the experiment to be reported below was designed with a view to testing these two competing hypotheses. The attempted replication was thereby embedded into a larger experiment which would seek to extend and clarify the nature of the effect. It consisted of a comparion between two groups of subjects, one group containing 13 pairs of non-close-relaters, and the other group 14 pairs of mere strangers. Note that both the relationship hypothesis and the personality hypothesis each predicted that non-close-relaters would display the significant sheep-goat correlation; the hypotheses differed only in their implications for the group of strangers. Note also that the highest sheep-goat correlations were expected to be observed in the percipients, though the sort of evidence which was presented in Part I of this chapter (namely, that from the Shuffle Controls, and the significant correlation observed for non-close-relater agents in the First Sheep-Goat study), led to the expectation that the agents as well would probably display significant correlations, though of a magnitude lesser than that for the percipients.

METHOD

Subject Selection. As a preliminary screening-device, the experimenter prepared a small questionnaire-booklet containing the

closeness-of-relationship question (see Appendix 2), plus the ten-item Sheep-Goat Scale. Because non-close-relaters are comparatively rare (less than one person in six), it was realized that the questionnaire booklet would have to be administered to a large number of subjects in order to find the desideratum of 30 volunteer non-close-relaters. Thus, during the period October 1978, the booklet was distributed at (1) a meeting of the student Psychological Society at Edinburgh University, (2) a student 'societies' fair', presented for the benefit of all freshmen entering that university, and (3) all classes of Psychology students at the university, that is, First through Third Year, with a sprinkling of Honours students and postgraduates. Of the 337 completed questionnaires, 50 respondents turned out to be non-close-relaters, and of these, 30 expressed willingness at the time to participate in an ESP experiment to be conducted later in the year. However, four of this number subsequently declined to take part, leaving a total of 26 non-close-relaters for the sheep-goat study; these persons were paired arbitrarily, except inasmuch as an attempt was made to match them 'across academic years', for example teaming a First-Year student with a Third-Year, to minimize the possibility of them being much acquainted prior to the ESP test. The same was done for the 28 persons selected arbitrarily to be in the 'mere stranger' group. All subjects were asked, upon arrival at the laboratory, to what extent they knew each other: most were complete strangers, and none fell into the category of being in an 'extremely close' relationship with their partner.

Characteristics of the Selected Sample. The 54 subjects comprised 28 males and 26 females, about equally divided between the two groups. Age at the time of the ESP test ranged from 17-36 years, with a median of 19.

Targets. Since the author's previous picture-guessing experiments had been conducted in Australia, there was generally no reason whatsoever not to use the target-sets already prepared in the case of this Scottish sample. (In a tiny number of cases, some of the subjects had carried out judging for the

experimenter, and were thus acquainted with a number of the potential targets; but since there was a record of which targets they had seen, the experimenter was able to ensure that different targets would be used to test the ESP-performance of that subject; but so as to avoid any possible methodological criticism, such subjects were in any case invariably designated the role of sender.) By this point in time, the experimenter had an additional six target-sets (prepared for the experiment to be reported in the next chapter, which chronologically preceded the present study), making a total of 18 sets, or 180 targets altogether. Since there were 27 dyads to be tested, some target-sets would have to be used more than once. In previous experiments (see, for example Chapter 3, pp 102), it had been so arranged that in a single test evening, just one target-set would be used, and from 1-4 pairs of subjects tested; in order to prevent sensory leakage, this procedure necessitated having all the evening's subjects arrive at virtually the same time (the waiting ones being cloistered away in a separate part of the Psychology Department), or at any rate in such a manner that no communication could occur between them as to the identity of the targets. In this present study, a somewhat different routine was adopted. The 18 target-sets were given to a colleague, who randomly selected three or four sets (without replacement), placing each inside a large, opaque manila envelope. The experimenter was given this pile of envelopes earlier in the day, and when it came time for the ESP-test, simply took the top one for use with the first pair to be tested that session, the next envelope for the second pair, and so on, subject only to the above-mentioned constraint that a target-set would be rejected if either of the subjects had previously seen it in the course of judging. When this pile was exhausted, the experimenter would obtain a new one from his colleague, containing sets randomly selected from those remaining in the latter's possession. When all 18 target-sets had been once used, they were returned to the colleague and the process repeated anew, until all 27 subject-pairs had been tested.* It

---

* The author wishes to thank Hilary Henegan and Kathy Wilson for performing this task.

might be objected that this procedure permitted subjects tested later in the series to have obtained some information from previously tested subjects about what the targets were; this is, of course, true, but the question is, how much useful information would have been so acquired ? Even if a percipient were perverse enough to have ferretted out, by large-scale detective work, each and every one of the 180 targets and the order of each within its own set of ten, he would still have to guess which of the 18 sets happened to have been selected for the occasion of his testing -- a long shot which, if he pulled it off, would be almost sufficiently unlikely to suggest the conclusion that the 'guess' was influenced by ESP! There was, in fact, no indication in the picture-guessing data that anyone had attempted to reproduce (from memory and inference) any target-set. In any case, the instructions to the receiver (see Appendix 6) led the subject to believe that any information such that he might have acquired from previously tested subjects would be thoroughly useless in the present session.

Procedure. All testing was carried out in the evenings, in the Psychology Department at the University of Edinburgh, the first pair on 6th December 1978, the 27th and last pair on February 8th, 1979. Pairs were tested individually, as usual. Upon arrival, subjects were welcomed and introduced to one another. They were then asked to complete a fresh copy of the (eleven-item) sheep-goat scale, with the explanation that it was necessary to know what their beliefs and attitudes were at that point in time rather than what they might have been some months previously upon initial testing. (As in the previous sheep-goat experiment, it was these sheep-goat scale-scores which were to be correlated with ESP-score, rather than those of the first test). As it happened, the test-retest reliability was reasonably good (although not quite so high as for the previous sheep-goat study, perhaps because here we were not dealing with extreme-attitude subjects): the retest interval ranged from 55-124 days, with a mean of 100 days, and for a sample of 54 the Sheep-Goat scale-scores correlated at $r = +.82$ ($2p = 6 \times 10^{-11}$). The ten scale-items gave test-retest correlations ranging from

rho = +.33 for Premonition to +.79 for Belief in Life After Death, and averaged +.57. Additionally, there were no significant differences between strangers and non-close-relaters on either the Sheep-Goat Scale as a whole, or on ten out of 11 of the component questions: the exception was Personal Experience of ESP, where 36% of strangers said that they had had such, compared to only 4% of non-close-relaters (χ^2 = 8.85, 2 df, p = 0.01). As to the responses given to the questionnaire by the 54 subjects as an undifferentiated group, 52% believed in the existence of ESP (only 6% did not), but only 4% would say that they were psychic, as opposed to 59% who said they were not; 43% claimed to have had at least one non-coincidental hunch, 24% a veridical premonition, 33% a precognitive dream, but only 4% a paranormal vision; 41% believed in life-after-death (44% being uncertain), and a somewhat smaller percentage believed that some people could make contact with spirits of the dead; 20% claimed an experience of telepathy, and the modal prediction as to how well they would perform in the ESP-test was the response "unlikely".

Subjects were next shown the original questionnaire-booklet which they had filled in in October, with a view to reviewing their responses to the close-relationship questions; so as not to pressure subjects into a stance they did not genuinely feel, it was emphasized that it was quite immaterial whether they had changed in the meantime since initial testing, and that what was important was their <u>current</u> status, whatever that happened to be. Of the 54 subjects, 48 maintained that their answers to the Closeness questionnaire were the same as they had been initially, while six said that they had changed: four of these were in the stranger group, and none had become non-close-relaters; however, two of these six were originally non-closerelaters, and each decided upon further reflection that they indeed really did have a close relationship with someone (thus making them close-relaters). Because of the difficulty of finding substitute non-close-relaters at so late a stage, these two subjects were nevertheless allocated to the non-close-relater group, but were assigned the role of agent

(it being expected that as far as the sheep-goat/ESP correlation was concerned, it was the percipient who would have the greater influence.) Aside from this rare occurrence, subjects were permitted to decide which role they preferred to take.

The ten-trial picture-guessing test then took place, preceded as usual by a single practice trial. During this practice trial, the experimenter placed the targets for that session into their cardboard windows (see Chapter 2, p. 47), in readiness for the test proper. It was thus only at this point in the procedure that the experimenter actually discovered which targets had been selected for use on that occasion; this eliminated the possibility of the experimenter inadvertently cueing the subjects during the pre-experimental interaction. It should further be pointed out that an additional precaution against sensory contact between agent and percipient was taken by separating them by a distance greater than had been possible in previous work (where adjacentrooms had been used, owing to the scarcity of laboratory space): on the present occasion, the two participants were seated in two small cubicles maximally distant from each other in a complex containing some 15 such cubicles, the intervening space measuring some 13 metres direct and 19 metres by corridor. As before, experimenter remained unobtrusively stationed in the agent's room in order to supervise the procedure where necessary. The doors of each of the two cubicles were kept closed during the test proper, and the only noise which reached the percipient had to be at shouting-level, and thus an auditory masking-device such as a fan was considered unnecessary. It might also be mentioned that, unlike the author's previous studies, in the present experiment agent and percipient were sat facing towards each other (albeit not face to face!), so as to facilitate the notion of the agent "communicating" with a person who was in the direct line of sight; in prior experiments using the adjacent-room laboratory-setup, agent and percipient were positioned such that they were side-by-side --- which some subjects reported as being psychologically disconcerting.

Immediately following the tenth and final picture-guessing trial, but prior to obtaining any knowledge as to their performance, each subject completed the standard post-experimental questionnaire (Appendix 7), in order to record how they had felt and how successful they thought they'd been etc. The percipient was then conducted to the agent's room for debriefing, as well as a comparison of the target-response pairs to see if there were any notable hits. The subjects were then thanked and escorted from the building.

Judging. A total of 54 persons participated in the rank-evaluation of the free-response data, most of the judges being unpaid volunteers from the sample of 337 students who were originally screened in the subject-selection process. The 27 sets of responses (10 responses from each of the 27 percipients) were ranked against each of the 10 respective targets aimed at by each percipient, three times and independently by different judges. Of these 270 'triple-evaluations', the inter-judge concordance for the three sets of ranks was significantly high on 51% of occasions (and thus non-significant on 49%), the mean Kendall's W coefficient of concordance being 0.63 for strangers and non-close-relaters combined and separately -- a highly consistent result.

RESULTS

I. The Post-Experimental Questionnaire.

In general, the responses given by non-close-relaters were very similar to those of the strangers, and the same may be said after comparing the responses of agents and percipients: 59% of all subjects were uncertain as to whether ESP had occurred, 33% believing it hadn't and only 8% saying it had. Forty-one percent suggested that there had been some factor present likely to inhibit ESP, the most frequently profferred theories being lack of acquaintance with their partner, the complexity of some of the target-stimuli (cited by agents), and the occasional occurrence of various

auditory or visual distractions. As regards the rating-scales for mood, and bearing in mind that the score-range is 1-6, the group as a whole reported themselves to be quite relaxed (mean = 4.4), very interested (4.7) and enthusiastic (4.6), but with a tendency to feel uncertain (3.7). On the 'calm-agitated' scale, however, there was a significant difference between agents and percipients, the former being more calm (1.9) than the latter (2.8) ($t = 3.0$, 52 df, 2p = 0.004). Again, significantly more agents than percipients (11% as opposed to 4%) considered that ESP had taken place, 56% of receivers denying this but only 11% of senders doing same ($\chi^2 = 12.13$, 2 df, $p = 0.002$). In summary, agents differed from percipients only in that they were more calm and more confident that ESP had taken place.

II. The ESP Test

(i) The Sheep-Goat Effect. The major hypothesis of interest was that the Sheep-Goat scale scores of the 13 non-close-relater percipients would correlate positively and significantly with ESP-score*. But quite contrary to prediction, the correlation was non-significantly negative or virtually zero: r = +.17 for mean rank, + 0.00 for hits. The story is similar for the 14 stranger percipients: - 0.00 for mean rank, -.27 for hits. Consequently, when these two subgroups of percipient are combined into one undifferentiated group, the correlation for the overall N of 27 is slightly negative rather than positive: r = +.04 for mean rank, -.14 for hits. So in this respect, then, the experiment was a very disappointing failure.

The situation is rather more optimistic, however, when we examine the results for the percipients' partners, namely the 27 agents. It was expected, with rather less confidence than for the percipients, that there could well be a significant correlation, and this is in fact what occurred: the sheep-goat scale-scores of the 27 agents correlated with the hits variable to the tune of r = +.48 (1p = 0.006), which is thus significant at better than the 1% level; similarly, the correlation with mean rank was -.42 (1p = 0.014).

---

* A one-tailed test of significance was thus appropriate.

Table 4.5 Correlations Between Number of Hits out of Ten, and the Sheep-Goat Scale and Component Items, Second Sheep-Goat Study.

	AGENTS			PERCIPIENTS		
	All A (n = 27)	NCR A (n = 13)	S A (n=14)	All P (n = 27)	NCR P (n = 13)	S P (n=14)
Sheep-Goat Scale 3	+.48**	+.47	+.49*	-.14	+.00	-.27
Belief in ESP	+.34*	+.29	+.40	+.02	+.22	-.22
Personal Experience of ESP	+.37*	+.28	+.37	-.11	+.07	-.35
Believe self to be Psychic	+.41*	+.28	+.46*	-.43*	-.68**	-.21
Have had a Hunch	+.12	-.37	+.52*	-.27	-.54*	-.02
" " " Premonition	+.02	-.26	+.30	-.06	-.09	-.05
" " " Precognitive Dream	+.32	+.42	+.18	-.04	+.45	-.36
" " " Vision	-.08	-.53*	+.35	-.04	+.44	-.46*
Belief in Life after Death	+.10	+.23	-.13	+.09	+.13	+.17
" " Spirit Contact	+.28	+.49*	+.12	+.03	+.24	-.14
Have had a Telepathic Experience	+.65†	+.85††	+.34	+.10	+.13	-.06
Estimated Probability of Success	+.44*	+.49*	+.36	-.17	-.33	+.08
Belief that ESP occurred	+.42*	+.44	+.50*	+.04	+.04	+.11
" " was inhibiting factor	+.16	+.27	+.05	-.36*	-.33	-.41

NOTE: * : $1p < 0.05$
** : $1p < 0.01$
*** : $1p < 0.001$
† : $1p = 0.00013$
†† : $1p = 0.00023$

'NCR' = 'Non-Close-Relater'
'S' = 'Stranger'
'A' = 'Agents'
'P' = 'Percipients'

Indeed, for the hits variable, 10 out of the 11 items comprising the Sheep-Goat Scale correlated positively with ESP, five of these ten doing so significantly (see Table 4.5, column one) by one tailed test: the most significant was Experience of Telepathy, which correlated r_s = +.65, (1p = 0.0001). For mean rank, all 11 items correlated positively with ESP, six being significant.

Now having obtained some highly suggestive evidence for a sheep-goat effect in the group of agents as a whole, the next task was to see whether or not this effect occurred to the same degree within the two subgroups -- non-close-relaters and strangers. The results seem to indicate reasonably unambiguously that the sheep-goat effect occurred within both groups to an equivalent extent: for the 13 non-close-relaters, the correlation (with total number of hits) was a marginally significant +.47 (1p = 0.052), and for the 14 strangers, an almost identical +.49 (1p = 0.036); for the mean rank variable, the correlation was -.49 for non-close-relaters (1p = 0.045), and -.39 for strangers (1p = 0.083). It therefore seems to be the case that the relationship hypothesis is to be preferred over the personality hypothesis: since strangers gave evidence for a sheep-goat effect comparable in quality to that given by non-close-relaters, it appears that there is nothing intrinsically special about the non-close-relater that is conducive to the effect; rather, it appears to be the case that the critical variable is the absence of a close emotional relationship between sender and receiver. We may thus feel reasonably confident in asserting that the experiment has yielded at least partial confirmation of an interactive sheep-goat effect; and the suggestion is that the effect is dependent upon the social situation between agent and percipient. This means that in future, if we want to pursue the sheep-goat effect, it will not be necessary to go to all the trouble of securing specifically non-close-relaters: all that must be done is to ensure that the two partners are relative strangers to one another.

(ii) Individual Scoring. The average rank-scores obtained by each subject-pair on each of the ten trials in the ESP test may be found in Appendices 14 and 15. Mean scores and other summary statistics for individual pairs are to be found in Tables 4.6 and 4.7. For the mean rank, the (inherently one-tailed) randomization-test probability must be less than or equal to 0.025, and none of the 27 pairs meet this criterion of significance; similarly, for the hits variable, in order to be significant, a pair must obtain fewer than 2 or in excess of 8 hits, but again none of the pairs achieved this level of scoring. However, two pairs of strangers appear to have obtained significantly higher variance, as measured by the 'mean absolute-deviation' score: whereas MCE is 1.90, pair 84 averaged 2.73 (1p = 0.005), and pair 95, 2.63 (1p = 0.019). In these pairs, therefore, there seems to have been a significant tendency to obtain highly extreme trial-scores (that is, either very high or very low), suggesting an alternation between psi-hitting and psi-missing over the course of the ten trials. It should be pointed out, however, that the binomial probability of obtaining two or more significant pairs out of 27, at the average probability of 0.012, is approximately 0.04, and multiplying this by 3 to correct for the three sets of analyses carried out (one for each of the three types of ESP score) would give us a figure in excess of 0.05, suggesting that the two significant pairs may simply be a result of multiple-analyses.

(iii) Overall Group Scoring. Using the technique of linear regression, and on the basis of the previously observed sheep-goat/ESP correlations, it was possible to construct an equation which could be used to predict the likely ESP-scores of a new sample of subjects. There is one equation for agents, and another for percipients, and one for each of the two ESP-scores (mean rank and hits), making four in total. On the basis of these equations, and assuming the stability of the previously observed correlations, it was predicted that the present group of strangers would score non-significantly

Table 4.6. Mean Ranks, Hit-Scores and Mean Absolute-Deviations for the Non-Close-Relaters (n = 13)

Pair	Average Rank			Number of hits	Absolute Deviation		
	Mean	S.D.	1p(Mean)		Mean	S.D.	1p(Mean)
83	4.30	1.98	0.054	8	2.13	0.90	0.355
79	4.53	2.15	0.086	6	1.97	1.29	0.326
99	5.07	1.85	0.259	8	1.50	1.17	0.732
94	5.23	2.09	0.344	6	1.67	1.28	0.825
90	5.30	2.72	0.426	6	2.40	1.29	0.113
77	5.50	1.80	0.486	6	1.47	1.05	0.826
78	5.63	1.69	0.437	4	1.40	0.96	0.853
96	5.90	2.65	0.312	4	2.33	1.32	0.213
80	6.00	2.43	0.228	4	2.23	1.07	0.075
101	6.03	2.50	0.255	3	2.17	1.36	0.431
76	6.20	2.58	0.149	3	2.33	1.31	0.110
88	6.50	1.39	0.064	3	1.50	0.83	0.796
98	6.53	2.33	0.078	3	2.20	1.29	0.274
Grand Total	5.595	0.70	-	4.923 (s.d. = 1.85)	1.946	0.38	-

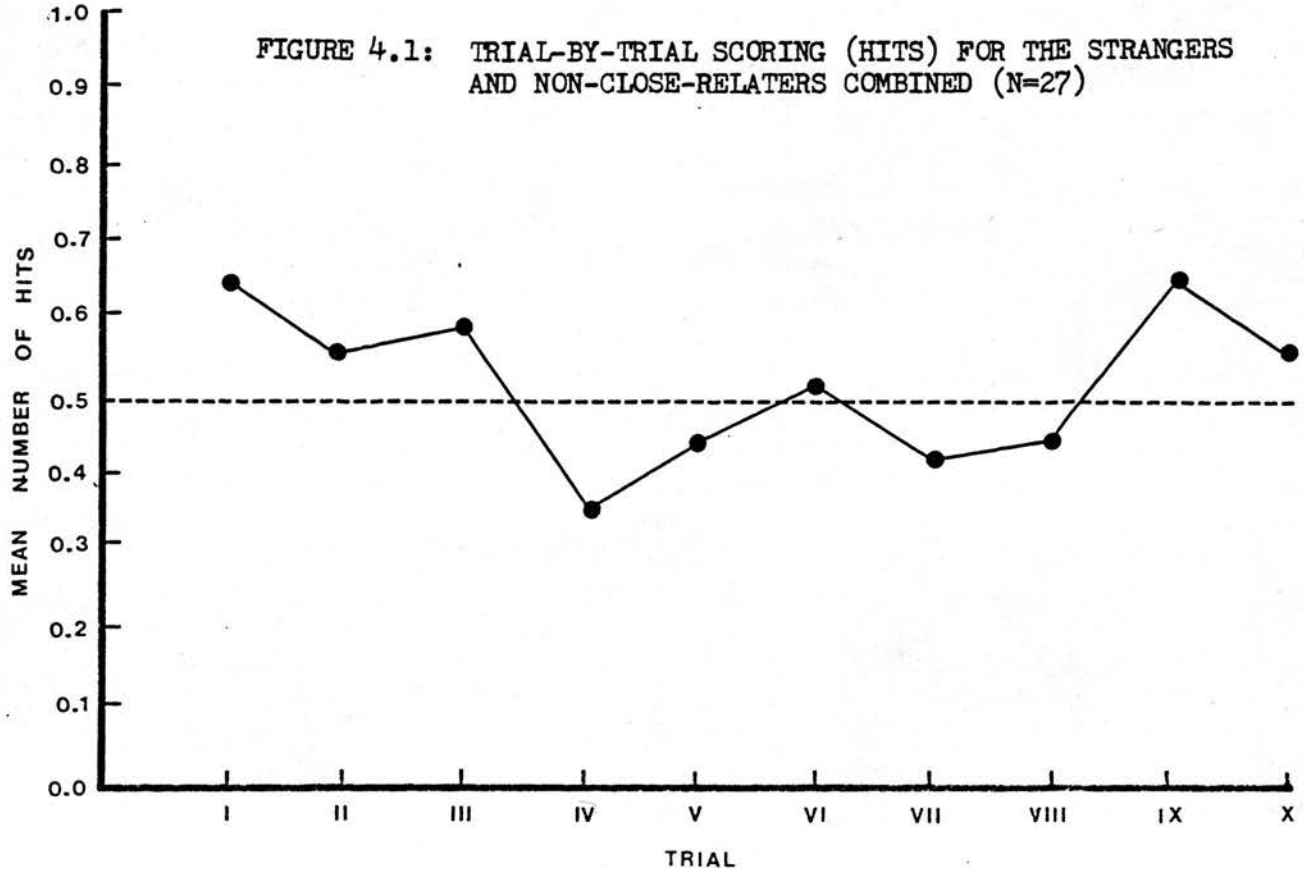
Table 4.7 Mean Ranks, Hit-Scores and Mean Absolute-Deviations for the Strangers (n=14)

	Average Rank			Number	Absolute Deviation		
Pair	Mean	S.D.	1p(Mean)	of hits	Mean	S.D.	1p(Mean)
97	4.67	1.83	0.119	8	1.77	0.96	0.727
89	4.70	2.23	0.140	6	2.07	1.16	0.389
93	4.83	2.06	0.172	5	1.83	1.15	0.664
85	4.90	1.63	0.211	7	1.43	0.98	0.927
100	4.97	1.84	0.216	6	1.47	1.23	0.833
92	5.10	2.61	0.304	5	2.50	0.84	0.089
87	5.17	2.11	0.318	8	1.60	1.41	0.869
91	5.33	2.47	0.400	5	2.20	1.13	0.188
82	5.37	2.06	0.436	4	1.47	1.45	0.736
84	5.63	2.98	0.403	6	2.73	1.18	0.005
95	5.67	2.82	0.435	4	2.63	1.03	0.019
81	5.83	1.78	0.329	2	1.47	1.06	0.874
102	5.93	2.43	0.283	4	2.03	1.40	0.347
86	6.40	1.66	0.109	4	1.43	1.23	0.890
Grand Total	5.321	0.52	-	5.286 (s.d. = 1.68)	1.902	0.47	-

higher than the non-close-relaters, but that the latter group would score significantly negatively. In the event, only the direction of the group-difference was correctly predicted, none of the grand mean ESP-scores deviating significantly from chance. (This failure is of course attributable to the fact that the sheep-goat/ESP correlation was not stable: it did not replicate at all for the percipients, and was somewhat weaker than expected for agents.) For mean rank, the stranger pairs' grand mean was 5.32, for non-close-relaters, 5.59 (t = 1.16, 25 df, 2p = 0.255); for hits, the stranger mean was 5.29, the non-close-relaters averaging 4.92 (t = 0.27, 2p = 0.792); for both types of score, the means of the strangers lay above chance, of the non-close-relaters, below. For the absolute-deviation measure (MCE = 1.90), strangers scored exactly at chance with a mean of 1.90, while non-close-relaters were above chance at 1.95 (t of the difference being 0.53, 2p = 0.598).

(iv) Serial-Position Effect. As in the previous Sheep-Goat experiment, the mean ESP-scores for each of the trials were subjected to repeated measures analysis of variance in an effort to determine (a) whether there was a significant change in level of scoring throughout the course of the test, and (b) whether the pattern of scoring conformed to a simple mathematical function, such as a straight line or a U-shaped curve. For each of these effects, a total of six analyses were made (strangers and non-close-relaters were examined separately and combined, for both the mean rank and the hits variable). None of these analyses indicated a significant tendency for level of scoring to change over time. However, trend analysis indicated that the combined group of 27 pairs, on the hits variable, displayed a significant U-shaped effect: F = 4.7045, 1 and 234 df, p = 0.0311. (A graph of the trial-to-trial scoring may be found in Figure 4.1).

FIGURE 4.1: TRIAL-BY-TRIAL SCORING (HITS) FOR THE STRANGERS AND NON-CLOSE-RELATERS COMBINED (N=27)



This significance might be regarded as spurious on over-analysis, were it not the case that the same ESP-measure yielded the same effect in the Closeness Experiment when the two major subject-groups (close-relaters and non-close-relaters) were likewise combined (see Chapter 3, p.109); though there was no such significant trend in the First Sheep-Goat Experiment, the result of pooling its data with those from the Closeness

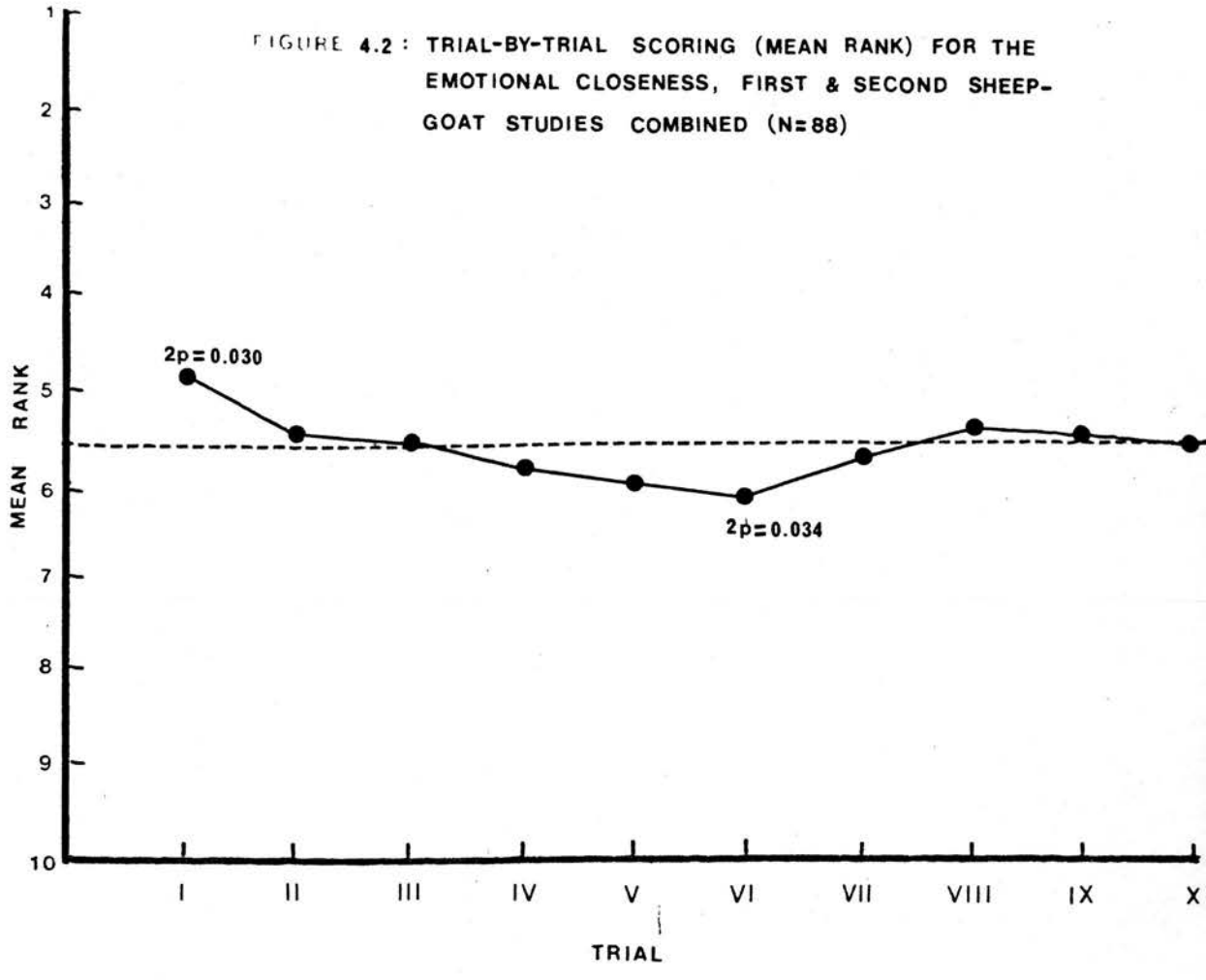


FIGURE 4.2 : TRIAL-BY-TRIAL SCORING (MEAN RANK) FOR THE EMOTIONAL CLOSENESS, FIRST & SECOND SHEEP-GOAT STUDIES COMBINED (N=88)

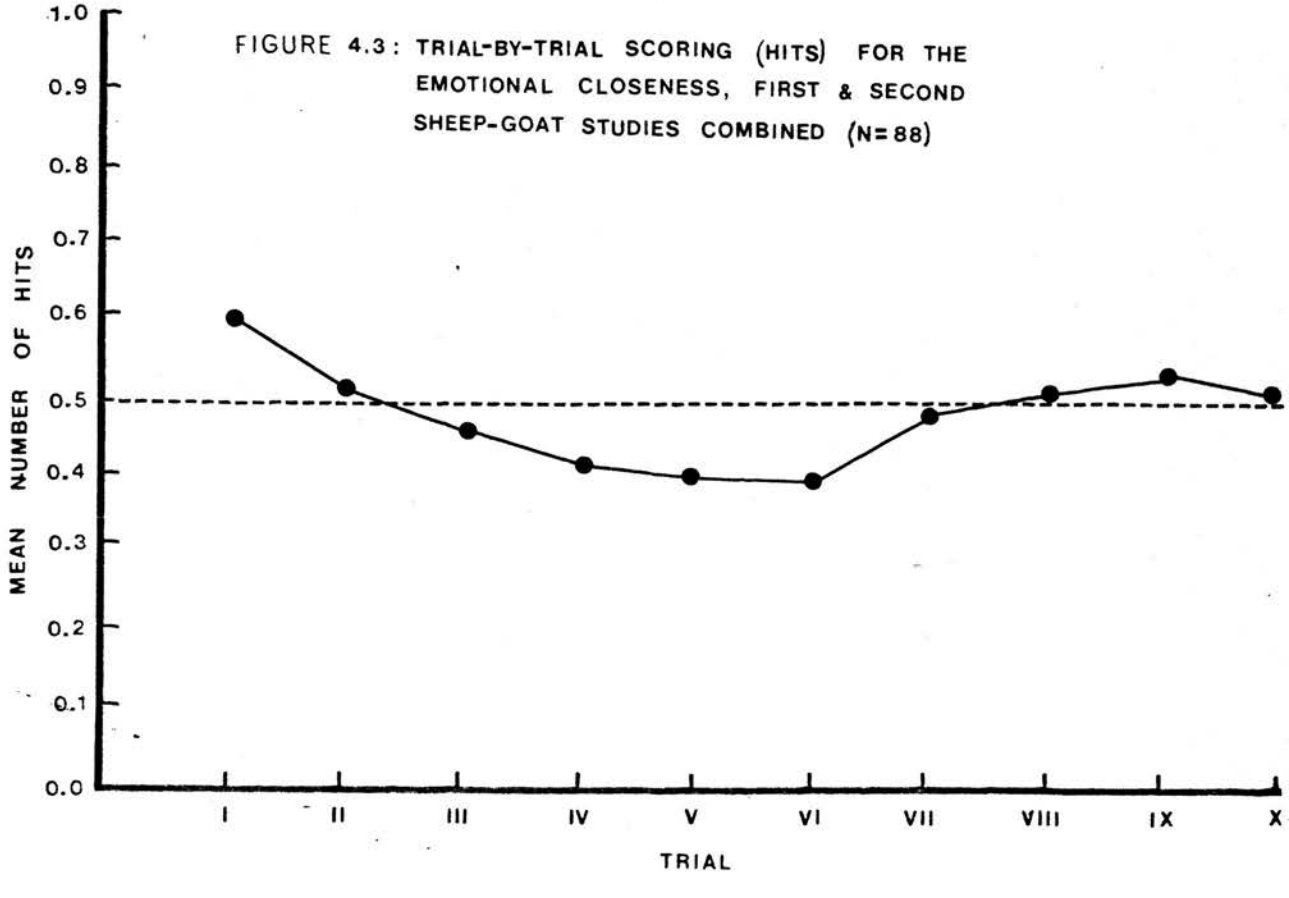


FIGURE 4.3 : TRIAL-BY-TRIAL SCORING (HITS) FOR THE EMOTIONAL CLOSENESS, FIRST & SECOND SHEEP-GOAT STUDIES COMBINED (N=88)

Study was to increase the significance of the effect (see Chapter 3, p.110). If now we combine the data from the three experiments thus far conducted, we find that for both the mean rank and the hits measure, each of the graphs of the trial-to-trial scoring for the combined sample of 88 subject-pairs does indeed resemble a shallow U-shaped curve (see Figures 4.2 and 4.3). Trend analysis confirms this visual impression: for the mean rank measure, the quadratic component's F-ratio for 1 and 783 df is 8.2426, p = 0.0042; for the hits measure, on this same component, F = 8.573, p = 0.0035. There thus seems to be a small but reasonably consistent tendency for the group-scoring in these ten-trial picture-guessing experiments to yield a U-shaped trend (provided that the sample-size is fairly large), that is to say, a pattern in which the scores tend to be higher at the beginning and end trials than in the middle.

Post hoc examination of the scores in Figures 4.2 and 4.3 suggests that the most effective division of the ten trial-means into high- and low-scoring groups is obtained by pooling the first three and last three trials together, and comparing them with the four intermediate trials (namely, trials 4 through 7): for example, the hits variable (MCE = 50%) yields an average rate of 52% on these six "end" trials, and 43% on the "middle" four (t = 2.75, 87 df, 2p = 0.007).

However, an alternative explanation for the observed curves is that what we are seeing is not a "fully symmetrical" U-shaped trend, but rather a decline effect, from first to sixth trials, followed by an upswing to what is essentially chance-scoring (rather than a return to psi-hitting). Evidence in favour of this proposition comes from two analyses: (1) if we perform

single-sample t-tests on each of the ten trial-means, for the mean rank score (though not for the hits variable), only the first trial is significantly above chance (2p = 0.03), while only the sixth is significantly below (2p = 0.034). On the U-curve hypothesis, one might expect a somewhat better showing on the tenth trial, yet the mean scores (5.515 for mean rank, 51% for hits) are very close to chance (see Appendices 19 and 20); (2) if we take the first six trial-means (that is, the apparent decline), and perform a repeated measures analysis of variance, the between-trials F-ratio for the mean rank is significant (F = 2.563, df = 5,435, p = 0.027), and moreover the linear trend is highly so (F = 12.329, df = 1,435, p = 0.0005). If, however, we take the last four (or five) trials (that is, the ostensible 'incline'), the differences between the means are non-significant, as is the test for linear trend. These two pieces of evidence thus combine to suggest that what we have is a significant linear decline from psi-hitting to psi-missing, followed by a regression to chance.

Be this as it may, the significant quadratic component is a very exciting finding indeed, all the more so because other free-response researchers have also recently reported such effects: Musso & Granero (1980), in a multi-session experiment where there were six trials per session (the single subject being Dr. José Muratti), also found a highly significant U-shaped trend (p = 0.006) in a subgroup of their data which had yielded overall evidence of psi-hitting. Again, Mary Makris (1979) carried out a ten-trial remote-viewing experiment (with trial-by-trial feedback), using eight non-close-relater percipients and eight close-relater percipients. Though the overall level

of ESP-scoring was significantly different between the two groups, their trial-to-trial scoring profiles were very similar, and each yielded a U-shaped curve, albeit an inverted one (for the combined N of 16, $F = 23.018$, 1 and 126 df, $p = 5 \times 10^{-6}$). We shall have more to say regarding serial-position effect in later chapters.

DISCUSSION

Though this experiment has yielded a number of hopeful signs of psi, it is nevertheless undeniable that it was to a certain extent a failure: to wit, a mammoth sheep-goat effect was expected for the percipients, and this prediction was completely without confirmation. What happened, then? The author feels very hard put to explain away the failed replication. There are, nevertheless, one or two small clues which might possibly offer some illumination. The first of these is the fact that in the routine check on the psychological equivalence of the two groups -- agents and percipients -- it was unexpectedly discovered that percipients reported themselves as being significantly more "agitated" during the ESP test than were the agents. Perhaps the sheep-goat effect also fails to occur in the presence of excessive anxiety; if so, then this would be very faintly reminiscent of the famous interactive sheep-goat effect reported by Schmeidler (1960), in which the effect occurred only for well-adjusted subjects and not for those showing signs of maladjustment on the Rorschach ink-blot test. (To cite another example indicating the fragility of the sheep-goat effect, let it be noted that Palmer & Miller (1972) found the effect to be eliminated in those groups where a monetary reward was offered for the highest score.) The second clue, related to the first, is that the two rating-scales used to measure relaxation and agitation during the ESP-test each correlated significantly with ESP-score for the 27 percipients to the tune of between .38 ($2p = 0.052$) to .48 ($2p = 0.012$):

in other words, percipients who were calm and relaxed tended to obtain higher scores than those who were tense and agitated. So perhaps, then, in the case of the percipients, anxiety was a more salient and determinative variable than their status as sheep or goats. And why might percipients in general feel more anxious than agents, when in all the author's previous experiments there has been no such difference? Perhaps this situation occurred because the spatial distance between agent and percipient was greater than it had been previously: instead of being in adjacent rooms, separated by a mere three metres, the participants in the present study were each in two rooms in a large complex of cubicles, distant from each other by about 13 metres as the crow flies. Perhaps such a distance can be expected to induce in subjects on the receiving end a greater sense of task-difficulty, which leads to apprehensiveness about their ability to succeed. It will be remembered, too, that percipients were significantly less confident than agents that ESP had in fact taken place. The obvious way of rectifying this problem in future research is either to bring the two percipients closer together, spatially, (without of course, allowing precautions against sensory leakage to be jeopardized), or alternatively to give more attention, as experimenter, to reassuring the percipient, alleviating any anxiety or diffidence they might feel, and generally helping them to view the task as being not insuperably difficult.

Despite the failure of the percipients to manifest a sheep-goat effect, the author feels sufficiently encouraged by the performance of the agents to suggest that the effect

should be pursued in future research. This attitude is engendered in part as a result of studying Table 4.8 below, which documents the 'progress', as it were, of the sheep-goat effect over the three experiments so far, separately and combined.

Table 4.8 Pearson correlations between ESP-score and the Sheep-Goat Scale, for non-close subjects only, First, Second and Third Studies separately and combined.

		MEAN RANK		HITS	
		r	2p	r	2p
Agents	1st (n=13)	-.68	0.010	+.57	0.040
	2nd (n=13)	-.54	0.057	+.29	0.328
	3rd (n=27)	-.42	0.029	+.48	0.012
	Combin. (n=53)	-.52	0.00006	+.49	0.00018
Percipients	1st (n=13)	-.79	0.001	+.80	0.001
	2nd (n=13)	-.69	0.009	+.45	0.121
	3rd (n=27)	+.04	0.880	-.14	0.500
	Combin. (n=53)	-.41	0.002	+.25	0.069

Note that the results of the Second Sheep-Goat Experiment seem to justify (i) pooling the non-close-relaters with the strangers, to represent the Third Study, and (ii) using, to represent the Second Study, the data reported in the footnote to page 122, rather than the data in Table 4.4.

Thus the table gives the sheep-goat/ESP correlations for all subjects who (without any doubt) did not have a close relationship with their partner.

One thing which is very striking about these results is the very evident trans-experimental decline in magnitude

of the correlations (the one exception being the upper right-hand quadrant, the correlations for agents with total number of hits). Some theorists, such as Millar (1979), would maintain that this decline is evidence that the psi-source is not the subjects themselves, but rather the experimenter, whose ability to cause, paranormally, his hypotheses to turn out as he wants them, is now on the wane. If this view is correct, then we would expect that if yet a third sheep-goat study were carried out by the author, the correlations would deteriorate even further towards chance (which seems already to have happened for the percipients). Only recently, we have been observing just such a phenomenon in the case of the Defense Mechanism Test (Haraldsson & Johnson, 1979) - a projective test which initially proved to be highly predictive of ESP (correlation of +.79 being reported by Carpenter, 1965), but whose predictive power has steadily declined to non-significance across eight experiments, this decline being so regular that the size of the DMT/ESP correlations itself correlates with chronological order at rho = +.93! Indeed, some would say that the trans-experimental decline is the most reliable phenomenon that we have in parapsychology!

It is, however, possible to take a different viewpoint, by supposing that the reason for the decline in size of correlation is that the later experiments did not sufficiently duplicate the conditions which together led to the sheep-goat effect in the initial study. We have already argued that the effect interacts with at least one other variable (namely, the degree of inter-personal intimacy between the subjects), and there

may well be others (such as level of anxiety). Another possible candidate for such a variable might be the degree of similarity between the sheep-goat attitudes of agent and percipient. In this Second Sheep-Goat Experiment, no attempt was made to pair participants who were 'like-minded': sometimes it so happened that a sheep agent was paired with a percipient who likewise was a sheep, but, as frequently, sheep agents ended up with goat percipients, and goat agents with sheep percipients. In fact, eight of the 27 pairs consisted of agent-percipient dyads characterized by 'harmonious' attitudes, while ten pairs of people had attitudes that were clearly dissonant, or 'antagonistic' (the remaining nine pairs not being readily classifiable, due to their non-extreme attitudes). It would be interesting to discover whether the sheep-goat effect is strongest when agent and percipient have similar attitudes; it would also be intriguing to know whether, when the attitudes conflicted, these 'opposing forces' cancelled each other out and produced no ESP whatsoever, or whether one side 'overpowered' the other in a battle of psychic wits, carrying the day by producing a level of scoring concordant with their own attitude. Preliminary and <u>post hoc</u> analysis of the data yielded by the present experiment suggests that the sheep-goat effect is slightly strengthened by having partners of consonant attitude: within the eight 'harmonious' pairs, the correlation between Sheep-Goat Scale and total number of hits is raised from +.48 to +.64 for the agents, and from -.14 to +.23 for percipients; however, when the partners 'dis-

agreed', what seems to have happened is that the agent determined the direction of the scoring: the correlation of +.48 rises to +.70 (2p = 0.024). These results must be considered merely as speculative pointers to future research, but nevertheless the author's own feeling is that some future research in this area is both justified and promising. For if we can take the results of Table 4.8 at face-value, and as indicating an effect that is occurring as a result of subject-psi rather than experimenter-psi, then it would seem valid to combine the data from the three experiments thus far conducted. We may then note with satisfaction that in the case of the 53 pooled agents, the Sheep-Goat Scale correlates with total number of hits to the extent of +.49, and with mean rank to a value of -.52; the odds against chance of a correlation of this magnitude being due to sampling error are, respectively, 5410 to one and 15,618 to one. Thus, chance seems an unlikely explanation for these results. We shall have more to say regarding the implications of these results in the discussion which forms Chapter 7.

CHAPTER 5

The Married Couple Experiment

CHAPTER 5

"Whom God hath put together . . . . ."
Matthew, XIX, 6

It seems to be a fairly widespread belief among persons who are married that the marital relationship often witnesses, and perhaps even helps to induce, the occurrence of ESP. If this belief is in fact correct, then presumably the cause of such paranormal events has to do with the deep emotional involvement of the two partners, in combination with their long-standing acquaintance with each other.

At least three published studies have provided evidence of psi-hitting in married couples. George Rice & Joyce Townsend (1962) compared a group containing three married couples and one engaged couple, with a group of four couples in which the partners had little or no acquaintance with each other. Each of the 16 subjects performed four runs of Zener cards under GESP conditions. Whereas mean chance expectation is an average of 5 correct guesses per run, the related couples averaged 6.50 ($t = 3.53$, 7 df, $2p = 0.0096$), the unrelated pairs 3.59 ($t = 6.23$, 7 df, $2p = 0.0004$), and there was a highly significant difference between the two groups ($t = 6.03$, 14 df, $2p = 0.00003$), by the present author's calculations. The experimenters also informally noted that the longer the couples had been related, the higher the score achieved.

Rice and his colleagues (1966) also studied the efficiency, for extra-sensory communication, of an intense emotion, with respect to the closeness of the emotional relations between agent and percipient. In one of their three experiments, agents were husbands, percipients their wives. Experimental husbands were subjected to immersion of their feet in ice-water, while control husbands experienced no such shock. Galvanic skin response (GSR) deviations were recorded for all wives, and it was found that experimental percipients yielded significantly higher GSR deflections than did the

control percipients. But while these findings do suggest that intense emotional stimulation of one person can affect another person by extrasensory means, the effect of being married is unclear, since there was no control group for the marital status variable.

Finally, Dolores Beer (1971) compared 15 married couples with 15 male-female pairs in which the partners had been strangers prior to the experiment. The task was a card-guessing one, and run under GESP conditions, each subject taking the role of agent once and of percipient once, as in the Rice & Townsend study. The unrelated group scored at chance, but the scores for the married couples were significantly high ($p = 0.001$).

There are at least two other studies, however, which have found no evidence of ESP in married couples. Brinkman & van Hilten (1972) used as target-material a series of short, tape-recorded episodes whose emotional tone was either positive, negative, or neutral. The agent listened to these tapes through ear-phones, and their spouse had to guess the emotional tone of the target. Non-significant results were found. James Carpenter (1977), in an unpublished study on the relationship between subject-rated mood of closeness between spouses and their GESP-scoring, found no evidence of a relation to mood.

One might also note parenthetically that there are other studies which have investigated the impact upon GESP of a close relationship between agent and percipient: examples are the successful experiment of Charles Stuart (1946), and the unsuccessful "sweet-heart" study conducted by John Beloff (1969) . These investigations are mildly relevant inasmuch as married couples comprised a percentage of the pairs constituting the 'close' group, but the data are not presented in such a way as to allow us to draw any conclusions from them about the effect of marital status per se.

Thus, the evidence for the proposition that the state of marriage represents a psi-conducive condition, while suggestive, is not without ambiguity,

and the present author hoped that the experiment to be reported in this chapter would help to strengthen the position. His reasons for supposing that significant results would be found were, firstly, that he still retained some faith in the Emotional Closeness hypothesis: though, as mentioned at the beginning of Chapter 3, the experiment which had been conducted to test this idea had not yielded a definitely significant difference between the 'close' and 'non-close' pairs, there were nonetheless marginal significances which suggested a real effect; encouragement was added to this belief by the observation that of the 31 pairs tested, the two who obtained individually significant results were the one married couple (who displayed high 'variance'), and the one engaged couple (who psi-missed). Though a sample of two is much too small from which to generalise, the significant married couple experiments were much more frequently reported in literature reviews than were the non-significant ones, and therefore seemed at the time to hold forth promise for the hypothesis. The author thus began to wonder whether the degree of closeness necessary to elicit psi-hitting might have to be of the intensity which leads two people to commit themselves to each other in a life-long partnership. Moreover, a free-response drawing experiment seemed eminently more suitable for testing the hypothesis than a forced-choice card-guessing technique, since the material involved in the former is more life-like and more interesting. The major hypothesis, therefore, was that the group of married couples as a whole would obtain an average score significantly above mean chance expectation. There was also a plethora of minor hypotheses tested in addition to this major one, and these will be detailed below in the course of the exposition.

METHOD

Subjects. It had been planned that a minimum of 15 couples would be tested, but due to sickness on the part of potential subjects and also to time limitations imposed upon the experimenter, only 14 pairs could be tested. This

might seem to be a rather small sample; but it could perhaps be argued that since it was possible for an individual subject-pair to obtain significant results (as per the Randomization Test), then surely it is not unreasonable to expect a sample of 14 to be able to yield evidence of an effect, especially since previous research had often been highly successful with small samples. If non-significant results were to be obtained from the group, then this would imply one of two things: either, the marital relationship has no effect at all on ESP, or, alternatively, while there may be an effect, it is very weak in general and therefore unlikely to be detected in a small sample. This experiment was thus on such a scale that only if the effect were fairly strong would it be likely to be observed. In addition, however, it was hoped to compare the results with those obtained from the earlier Closeness Experiment. We would then have three groups located on different points of the closeness continuum: one group where the people were relative strangers, a second group where the people had a close relationship with each other, and a third where all the subjects were in fact married. So the married couple experiment can be considered an extension of the earlier study, the latter thus providing the control group.

For reasons which seemed good at the time, the experimenter decided to limit the study to couples who had been married for less than ten years. This was done partially with the idea in mind that a test might eventually be made of a group of long-married couples (that is, people who had been wedded longer than (say) 20 years). But it was also the author's belief that people would feel closer to one another in the first few years of their marriage than they would after they had been together for a good number of years (although some readers may well wish to question this assumption), and he was inclined to consider sheer emotional closeness as being the factor most likely to be conducive to psi-hitting. Nevertheless, it was planned to test the hypothesis that the longer the couple had known each other, the better would be their ESP-score, on the grounds that sheer familiarity might

be an important variable; to this end, note was made as to how long the couple had been married prior to the experiment, and also how long they had known each other altogether. Length of acquaintance ranged from 13 months to nearly 10½ years, with an average of nearly 5 years. Length of marriage averaged just over 3 years, and ranged from 4 months to 8 years. Subjects ranged in age from late teens to early thirties. They were mostly friends and acquaintances of the experimenter, and came from a very broad spectrum of occupational and educational backgrounds.

Targets. Since a number of subjects were colleagues of the experimenter who had seen many of the targets already prepared at the time, it was deemed desirable to select an entirely new set for use in the current study. Using the procedure described in Chapter 2, sixty new targets were chosen and depicted, and grouped into six sets of ten. Great care was taken to ensure that the targets were as good as the experimenter's artistic ability permitted, since in previous studies an occasional agent would remark that representational inaccuracies or ambiguities would distract them from the task of concentrating. These sets were numbered sets 13 to 18, and are listed in Appendix 1.

Procedure. The experiment was carried out in the fortnight between August 17th and 30th, 1977, in the Psychology Department at the University of Adelaide. Upon arrival at the laboratory, and after the taking of routine recording-data, each subject was asked to complete the 11-item sheep-goat inventory. (This experiment chronologically succeeded the First Sheep-Goat Experiment (Chapter 3), but pre-dated the Second (Chapter 4), and therefore it had not been known at the time that the Sheep-Goat Effect was unlikely to occur if the partners possessed a close relationship.) Then followed the standard ten-trial drawing-reproduction test, the roles of agent or percipient being allotted according to subject-preference. The laboratory set-up consisted of the two adjacent rooms described in Chapter 2 (pp. 48-50.), the

percipient being seated at a table in the inner room, while agent and experimenter were stationed in the outer room, the adjoining door being closed and the fan running all the time. At the conclusion of the tenth trial, subjects were administered the standard post-experimental questionnaire (Appendix 7).

It should be mentioned that the target-sets were used in the same sort of sequence as for the First Sheep-Goat Experiment (See Chapter 3, p. 102): that is, one set would be used per evening session, with a maximum of four subject-pairs being tested (consecutively) in the same evening, their arrival and departure at the laboratory complex being judiciously timed so that the earlier-tested pairs would not come into contact with pairs that had yet to be tested in the same session.

The final part of the testing-session concerned two ideas which we may refer to as the Participation hypothesis and the Warcollier hypothesis, respectively. The first idea, which is probably attributable to Stuart (1946), is based on the notion that in a free-response experiment, the response may be related to the target, but in such an obscure way that the relationship is difficult to recognise. From this it might follow that the best detectors of these obscure relationships would be not independent judges but the agent and percipient. These three categories of participant have different sets of experiences within which to select what they consider to be salient relationships: the agent has not only the targets but his associated experience while looking at each target, to compare with a limited number of expressions and meanings that the percipient has been able to convey in their responses; the percipient has the advantage in that he is comparing the target (which may be complex but is not obscure or uninterpretable, as a response may be) to a response which he is able to understand and interpret better than any observer; the independent judge has the task of educing resemblances between the target and response without the benefit of personal interpretation of either. From this follows the Participation hypothesis, that is, the idea that the persons who participated in the experiment may prove

to be better judges of their own data than are independent judges.

At the same time, it would seem that a similar consequence is derivable from what we may call the Warcollier Hypothesis. René Warcollier, it will be remembered from Chapter 1, was a famous French parapsychologist who did much interesting work on drawing-reproduction during the 1920's and '30s. Though he believed that a "preliminary emotional connection" between agent and percipient was necessary for the occurrence of telepathy, he seemed to think that a very close relationship actually hindered transmission of the target-information. He speculated that this was because contemplation of the target-drawing aroused in the sender a whole chain of associations, and that it is these associations (rather than the specific idea inherent in the target) that are more likely to be aroused in a percipient who shares a common set of experiences. Thus, a close relationship does not inhibit telepathy per se, but rather, adversely affects transmission of the target-drawing.* Now an implication of this phenomenon is that some of the target-response resemblances may be rather idiosyncratic to that particular agent-percipient pair, and due to their esoteric nature may be unrecognizable to an independent judge. It thus follows, from both the Participation hypothesis and the Warcollier hypothesis, that agent and percipient should each do better at judging their own data than do independent judges. Though the married couple might not be able to obtain scores significantly in excess of chance, they might nevertheless perform significantly better than non-participant evaluators.

---

* A possible instance of this phenomenon may actually have occurred in this Married Couple Experiment when, on one occasion, the target was a pram, and the percipient drew a racing-car: the agent reported that while he was looking at the target-drawing, the pram brought to mind a friend whose wife had just had a baby, and the fact that this friend was a racing-car driver; it is thus, apparently, the association that got across to his spouse.

In order to test these hypotheses, use was made of the so-called Method of Correct Matchings (see Chapter 2, pp.58-60). The Preferential Matching Method was not used in this part of the procedure because it was considered rather too time-consuming, and more particularly so as to avoid jeopardising the Assumption of Independence. (Bias may be introduced when the same pool of pictures is repeatedly ranked, even though against different master-pictures, by the same evaluator: see Chapter 2, pp. 63-64). To refresh the reader's memory, let it be pointed out that in the Method of Correct Matchings, all ten targets are laid out in randomized order, together with all ten response-drawings, likewise randomized, and the judge is asked to pair off each response with its own proper target. The response is either correctly matched to its own target, (in which case it is scored as a 'hit' or 'correct match'), or it is matched, incorrectly, with one of the other nine targets (in which case it is deemed a 'miss' or 'incorrect match'); the range of possible scores under these circumstances is therefore 0 to 10 (excluding 9).*

Judging. (i) Preferential Matching. The 14 sets of response-drawings were preferentially ranked for degree of correspondence to each of their ten

---

*An unfortunate source of possible sensory cueing exists for the agent inasmuch as, when he knows the percipient well, he may be able to estimate which response-idea was likely to have been the first or last, or to have had some recognizable serial relation to the others in the sequence of ten guesses; this, combined with his memory of the target-order, may lead to a spuriously high number of correct matchings. Fortunately, there is no such cue for the percipient: the targets were originally chosen by a random process, and therefore display no seriality cues, and thus the memory of his own response-sequence will be of no avail. The seriality problem was minimized by randomizing the target-order and response-order before presentation to any of the judges; this seems to have seriously disturbed, if not completely destroyed, the agent's memory of the target-sequence. However, this could not prevent the possibility that an unscrupulous participant might have marked each of his ten pictures (targets or responses) with an unobtrusive code revealing its ordinal position to his or her partner; ideally, duplicate target-sets would be used for the correct-matching evaluation — a practice which was instituted after ten pairs had already been tested (prompted by practical considerations rather than by any suspicion that fraud was in fact occurring.) But as it happens, the possibility of spurious scores is relatively academic: the agents' scores were higher than those of the percipients and the independent judges, but not significantly higher; indeed the number of correct matches obtained by agents was non-significantly below chance!.

relevant targets, by a total of 35 persons, resulting in three independent rank-orderings for each response-set in relation to each target. For the resulting total of 140 triple-evaluations, Kendall's W coefficient of concordance averaged out at 0.63; it was significantly high (i.e. equalled or exceeded 0.627) in 59% of all cases, and was thus non-significant (i.e. fell short of 0.627) on 41% of occasions.

(ii) Method of Correct Matchings. Random number tables were used to obtain a randomized order for each target-set and for each response-set, and the two sets were laid out adjacently on a long table for presentation to the judge.* The participant judges were each instructed in the method they were to adopt in pairing the targets and responses, and were given a brief resumé of the sorts of resemblances to look for (that is, similarities based upon form or association, or even upon more personal and idiosyncratic relations such as the pram/racing-car example.) Each subject performed their matching-evaluation independently of the other (who waited, with the experimenter, in an adjacent room), and communication between the two was discouraged until both had finished their evaluations. When the target-response matchings made by both evaluators had been recorded (on a special record-sheet constructed for this purpose: see Appendix 11), the experimenter revealed the true pairings, and the results were discussed before the subjects took their leave.

The independent judging was performed twelve months later, by four male Psychology students at the University of Iceland, Reykjavik.** Two of these persons evaluated the data of seven couples (viz. those whose identity numbers were odd), while the other two evaluated the data of the seven

---

* It will be remembered that each target and each response had, prior to the test, been inscribed with a randomly-selected four digit code-number, and thus revealed no clue as to its ordinal position in the sequence of ten.

** I would like to express my gratitude here to these four students for undertaking this labour: Lúðvíg Lárusson, Sæmundur Hafsteinsson, Bjarni Ingvarsson and Halldór Bjarnason.

even-numbered pairs, thus providing, for each pair, two separate and independent non-participant evaluations. Consequently, for each of the 14 agent-percipient pairs, there are four 'correct-matching scores', namely the total number of correct matches obtained by i) agent, ii) percipient, iii) first independent judge and iv) second independent judge.

At the same time, it was of subsidiary interest to enquire as to what degree of concordance existed between the four evaluators; that is, irrespective of how many correct matches were made, to what extent did the judges agree that (say) response **x** should be paired with target **y** ? To the author's knowledge, no statistical work has ever been done to answer this sort of question, and he therefore proceeded to approach the problem in the following way. The reader is referred to Table 5.1, where A through D stand for four different targets, and A' through D' their four respective responses.

Table 5.1

Schematic representation of a four-trial experiment showing correct-matching evaluation and inter-judge concordance

		EVALUATOR				
Target	Response	Agent	Percip.	Ind.J.1.	Ind.J.2.	4-Judge Concord.
A	A'	A'	A'	A'	D'	3/4 → 0.75
B	B'	C'	D'	C'	B'	2/4, 1/4,1/4 → 0.25
C	C'	D'	C'	D'	C'	2/4,2/4 → 0.50
D	D'	B'	B'	B'	A'	3/4 → 0.75
Σ correct:		1	2	1	2	Mean = 0.56

The middle four columns record which particular responses were paired up with targets A to D in the course of the Method of Correct Matchings, by, respectively, agent, percipient, first independent judge and second independent judge. One compares each of these columns with the target-order in column one so as to obtain the score denoting the number of of correct matchings obtained by each evaluator. For instance, the percipient correctly matched response A' to target A and response C' to target

C, and is therefore awarded a score of '2'. Reading horizontally, however, and disregarding whether the response is correctly matched to its own proper target, we can award a score that reflects the number of judges who matched the same response to the same target. For example, in row four, three out of four of the judges agreed that response B' was the best match for target D. On the analogy of a Kendall's W, we may score unanimous agreement as 1.0, complete disagreement as 0; for a quadruple-evaluator situation, there are only three additional and intermediary degrees of concordance, and these may be scored between 1 and 0 at equal intervals: thus, if 3 out of 4 judges agree (only one dissenting), as in row one, let us score it 0.75; with two judges agreeing on one choice and the other two agreeing on a different choice (a divided house, as in row three), give 0.50; and when two judges agree but each of the other two have two different choices (as in row two), give a somewhat lower score, namely 0.25. Thus, successive grades of agreement are separated from each other by 0.25 points (which represents unity divided by the four possible states of inter-judge discordance.) We may then find the average concordance score for all the rows in the table (in our hypothetical example, 0.56), and then, too, for a number of tables, which is to say, a group of subjects. In this manner, then, was a four-judge concordance-score assigned to each of the ten horizontal rows of each of the 14 pairs in the married couple study. For the group as a whole (that is, for a total of 140 rows), the mean concordance score was 0.39, which indicates that on average, within each row, at least two of the four evaluators would agree as to which of the ten responses best matched the target in question.

A somewhat different way of measuring inter-judge agreement (which nevertheless correlates almost perfectly with the above-described way) is to take all possible pairs of evaluators, and for each pair, simply count up the total number of matchings agreed upon by the two (a score which obviously

ranges from 0 to 10, excluding 9). These totals may then be tabulated in a 'correlation matrix' such as in Figure 5.1 below.

Figure 5.1

'Correlation matrices' for pairwise agreement between judges, (left) for fictitious data in Table 5.1, (right) for married couple data.

	Ag.	Per.	IJ1	IJ2
Ag.		2	4	0
Per.	2		2	1
IJ1	4	2		0
IJ2	0	1	0	

Mean = $\frac{9}{6}$ = 1.50

	Ag.	Per.	IJ1	IJ2
Ag.		2.8	3.1	2.6
Per.	2.8		2.9	2.6
IJ1	3.1	2.9		3.6
IJ2	2.6	2.6	3.6	

Mean = 2.90

The left-hand side of the figure shows that, for example, the agent and first independent judge (IJ1) agreed on four matchings, while first and second independent judges agreed on none. The right-hand matrix shows the mean agreement totals for the married couple data: for instance, the two independent judges agreed, on average, concerning 3.6 target-response pairings out of a possible ten. For the six evaluator pairs as a whole, the group mean was 2.90, which indicates that if we take any pair of the four evaluators, they would be most likely to have agreed on nearly three of their ten matchings. Repeated measures analysis of variance, applied to the six group means, yielded non-significant results. The conclusion to be drawn from this is that the degree of concordance between pairwise combinations of evaluators was relatively consistent around 2.90. Thus, for example, the two independent evaluators did not agree with each other (mean = 3.6) to a significantly greater extent than did (say) independent judge two and the agent (mean of 2.6).

RESULTS

1. The Pre- and Post-Experimental Questionnaire

Examination of the sheep-goat questionnaire responses given by the 28 subjects revealed no significant differences between agents and percipients, or between husbands and wives. Speaking, then, of the sample as a whole, we may record that 54% believed in the existence of ESP (the remaining 46% being uncertain), while 32% believed they'd had personal experience of ESP, with just 7% claiming to be psychic (57% denying that they were thus); 43% claimed at least one non-coincidental hunch and one veridical premonition, while a somewhat smaller number (25%) claimed experiencing a precognitive dream, though only 11% had had a paranormal vision. Just over a third of the sample believed in life-after-death, but only about half that number thought contact with spirits of the deceased was possible; 36% claimed to have had an experience of telepathy.* The modal response given to the question as to how likely the subject deemed it that they would display ESP, was "50:50".

As regards the results of the post-experimental questionnaire, 82% were uncertain as to whether ESP had in fact occurred, while 14% thought that it had and 4% not. More than one third of the sample suggested that there was at least one factor present that could have lessened the likelihood of ESP occurring, the most frequently cited factors being fatigue (and consequent difficulty in concentrating) and the noise of the fan employed to provide an

---

* On the suspicion that some of these percentages were slightly more elevated than is usually observed for unselected subjects, the author re-examined all the sheep-goat questionnaire data currently to hand, comparing the responses given by married and by non-married persons (the former group including divorcees and widowed subjects). The 28 married persons in this current study were reasonably comparable to the 77 married subjects tested in other (psychological) experiments. However, when this group of 105 married persons was compared to the 744 single persons, the former reported significantly more personal experience of ESP (particularly premonitions and telepathy) and more frequently claimed to be psychic. But, as John Beloff points out (personal communication), age may be a contaminating variable: the older one becomes, irrespective of being married, the more likely one is to experience psi or psi-like coincidences. In fact, the technique of partial correlation subsequently demonstrated that when the effect of age was controlled for, the correlation between experience of psi and marital status disappeared.

auditory masking-screen. As far as the mood-scales, the mean scores were very comparable to those found in previous studies: bearing in mind that each scale ranges 1-6, the group as a whole reported themselves to be quite relaxed (mean of 4.4), extremely interested (5.2) and enthusiastic (5.0), quite low on agitation (1.7), though middling with respect to degree of confidence (3.4).

II. The ESP Test

(i) Group Scoring. The major hypothesis under consideration was that the grand mean ESP scores would deviate significantly from chance in the positive direction. However, no significant departures from MCE were obtained. The grand mean rank of 5.61 was non-significantly below chance of 5.50; the mean total number of hits exceeded chance of 5, but was only 5.14; and the grand mean absolute-deviation of 1.89 was below chance of 1.90. The scoring based on the Method of Correct Matchings likewise failed to detect any evidence of psi. Whereas for 14 sets of responses the expected total number of correct matches is 14 (with variance the same), the 14 agents achieved a total of 12 hits, percipients 9, and each of the two sets of independent judges exactly 7; obviously none of these totals exceeds 14, and indeed the independent judges' score of 7 is almost significantly negative ($z = 1.871$, $2p = 0.06$)! Repeated measures analysis of variance revealed no significant differences between the means of these four sets of evaluation. Again, while the participant-judges (i.e. agents plus percipients combined) on average obtained a higher number of correct matches than did the non-participant evaluators (i.e. the independent judges, combined), the means being 10.5 and 7.00 hits respectively, a t-test for paired samples indicated that the difference was insufficiently great to be significant ($t = 1.24$, 13 df, $1p = 0.12$). That is to say, the married couples were

not really much better at judging their own data than were independent evaluators. Support of only the weakest kind (namely, confirmation of the predicted direction of difference) has thus been obtained for the Participation and/or Warcollier hypotheses.

(ii) Other Miscellaneous Hypotheses Pertaining to Group Scoring

(a) Effect of Length of Marriage. It was predicted that there would be a positive correlation between ESP-score and length of marriage (and/or acquaintance). But, contrary to prediction, the direction of the correlation was opposite to that predicted: the correlation between total number of hits and time married was -0.40, ($2p = 0.153$), and for mean rank was +0.53 ($2p = 0.052$). Thus, there was an almost-significant tendency for those persons married for a longer time to score more negatively than relatively newly wed couples!

(b) The Sheep-Goat Effect. The sheep-goat scale-scores were correlated with ESP-score, in the hope of finding evidence of sheep scoring positively and goats negatively. For the 14 agents, the correlations were non-significantly negative, while those of the 14 percipients were positive but remarkably close to zero. The most that can be said for this null finding is that it is at least consistent with the results of the sheep-goat experiments reported in the two preceeding chapters, which suggested that the sheep-goat effect would not be found unless agent and percipient were relatively unacquainted with each other. Thus, though the present failure to reject the null hypothesis cannot of itself be taken as evidence for psi, the fact is at least consistent with the model proposed for predicting the occurrence of the sheep-goat effect.

(c) The Sybo Schouten Hypothesis. The Dutch researcher Sybo Schouten has recently (1979) published a report of a computer-analysis of 562 well-attested cases of ostensible spontaneous telepathy, selected from the

collection published in Phantasms of the Living (Gurney, Myers and Podmore, 1886). Schouten's aim was not to prove the existence of ESP, but rather to study whether or not non-psi factors (such as reportage artifacts) could explain the relationships which seem to hold between certain variables, and if not, to generate hypotheses under the assumption that ESP had occurred in the cases. One of the hypotheses proposed by Schouten was that "within a relationship characterized by mutual emotional ties, the probability of becoming a percipient is higher for the partner who is more emotionally dependent, while the probability of becoming the target-person is higher for the partner who is the object of this emotional dependency" (pp. 425-426). Part of the evidence that led Schouten to formulate this hypothesis was the observation that in cases of telepathy between spouses, the percipient was the wife in nearly twice as many cases as was the husband. The present study thus provided perhaps one of the first opportunities to test this conjecture in an experimental context. It is of course desirable to determine whether in fact the majority of modern marriages are characterized by greater emotional dependency on the part of the wife, but if we can assume that this is so, then Schouten's hypothesis might predict that the mean ESP-score would be higher when the wife is percipient than when the husband takes this role. Of the 14 pairs tested altogether in the present experiment, the wife was percipient in five, the husband receiver in nine. The difference between the ESP-scores was in fact in the direction expected: for the mean rank, the wife-percipient mean was above chance at 5.49, the husband-percipient mean below, at 5.68 (t-ratio of the difference = 0.75, 12 df, lp = 0.23); likewise for the hits variable, wife-percipients scored non-significantly above chance with a mean of 5.80 hits out of ten, husband-percipients non-significantly below, averaging 4.78 (t of the difference = 1.48, lp = 0.08). While neither of the differences is significant, the second is

Table 5.2 Mean Ranks, Hit-Scores, and Mean Absolute-Deviations for the Married Couples (n = 14)

Pair	Average Rank			Number of hits	Absolute Deviation		
	Mean	S.D.	1p(Mean)		Mean	S.D.	1p(Mean)
63	4.93	2.01	0.209	7	1.60	1.34	0.721
71	4.93	1.99	0.203	7	1.73	1.13	0.692
74	4.93	2.62	0.248	6	2.30	1.38	0.200
69	5.27	1.87	0.366	7	1.47	1.19	0.754
73	5.50	2.29	0.479	4	2.07	0.99	0.642
62	5.60	1.92	0.492	6	1.57	1.11	0.886
67	5.63	2.54	0.433	5	2.10	1.44	0.377
64	5.73	1.79	0.333	5	1.47	1.06	0.829
66	5.73	2.18	0.371	5	1.87	1.15	0.659
68	5.90	2.02	0.279	4	1.73	1.12	0.669
75	5.93	2.61	0.283	4	2.30	1.31	0.190
65	6.13	2.19	0.195	5	2.00	1.09	0.387
70	6.13	2.31	0.203	3	2.07	1.21	0.397
72	6.17	2.27	0.170	4	2.13	1.02	0.335
Grand Total	5.610	0.45	-	5.143 (s.d. = 1.29)	1.886	0.29	-

approaching the 5% level, and may perhaps be considered encouraging in view of the rather small sample.

(iii) Individual Scoring. The average rank-scores obtained by each of the 14 subject-pairs on each of the ten trials in the ESP-test, may be found in Appendix 16. Mean scores and other summary statistics for individual pairs are to be found in Table 5.2, as usual the pairs being listed in order of magnitude of their mean rank. In the original Closeness Experiment, the sole married couple and the sole engaged pair were the only pairs to obtain individually significant results. Encouraged by this outcome, it was expected that a significant percentage of the 14 married couples would likewise display significant results. But disappointingly, as detailed inspection of Table 5.2 will reveal, not one of the pairs obtained scores which were significantly different from chance, on any of the three measures of ESP.

(iv) Serial-Position Effect. Following the promising results obtained in the three previous studies, the hypothesis was advanced that here in the Married Couple Experiment there would occur evidence of a serial-position effect; that is, there was expected to be a significant change in scoring-level over the ten trials in the ESP-test, or (more likely) the appearance of some mathematical pattern, such as the U-shaped curve. The grand mean rank-scores, and the mean number of hits, per trial, are displayed graphically in Figures 5.2 and 5.3 respectively. The 'change-in-level-of-scoring' hypothesis was tested using the between-trials F-statistic yielded by repeated measures analysis of variance, but no support was obtained, despite the apparent incline from low to high scores over the course of the test. Nevertheless, partitioning of the Sum-of-squares into the various trend components yielded a significant linear trend: for mean rank, $F = 6.007$, and with 1 and 117 df, $p = 0.016$, while for hits, $F = 8.154$, $p = 0.005$. This is the only clearly significant result found in the entire experiment, and in view of the large number of statistical analyses carried out, must be regarded with a caution

FIGURE 5.2 TRIAL-BY-TRIAL SCORING (MEAN RANK) FOR THE EXPERIMENT WITH MARRIED COUPLES (n=14)

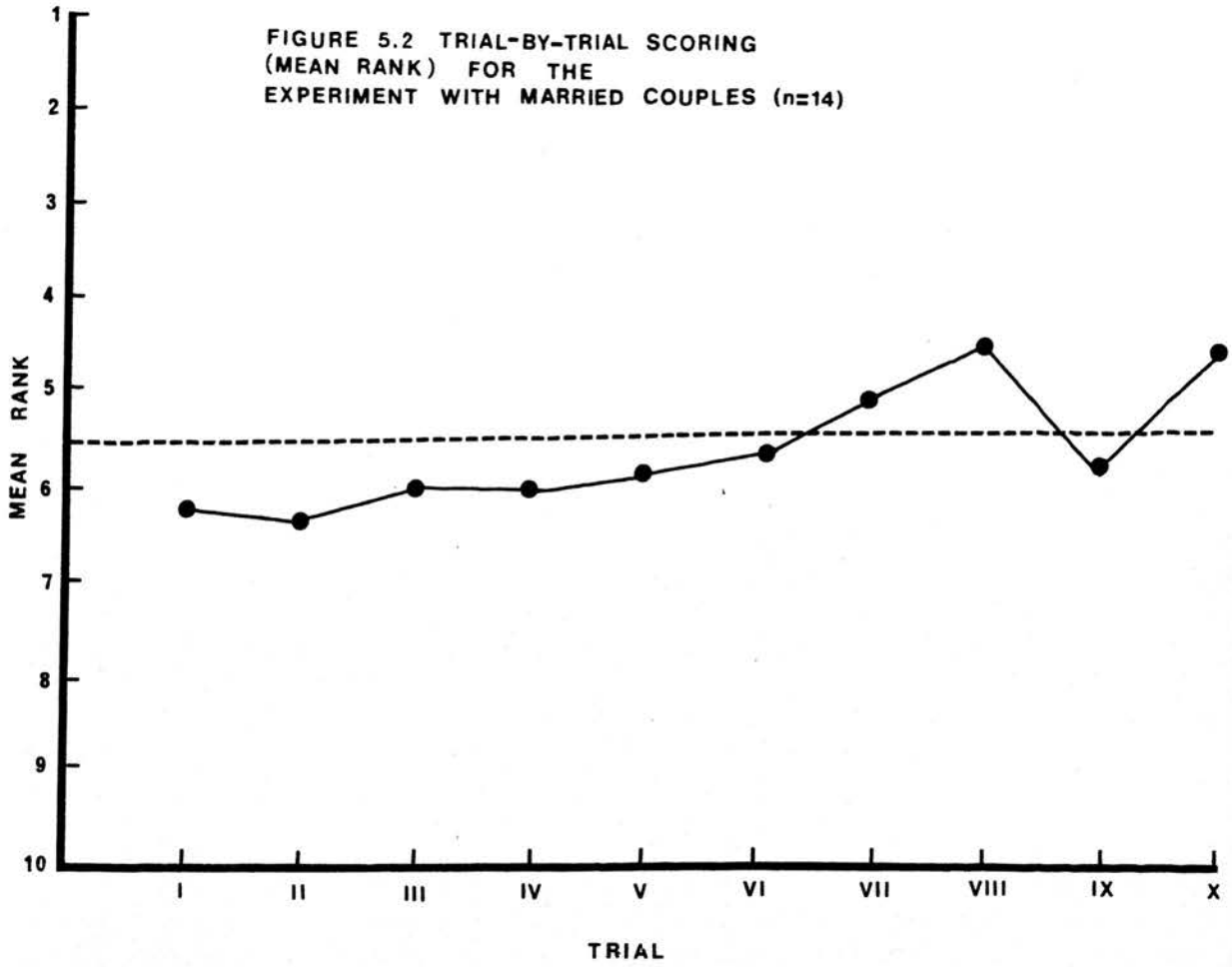
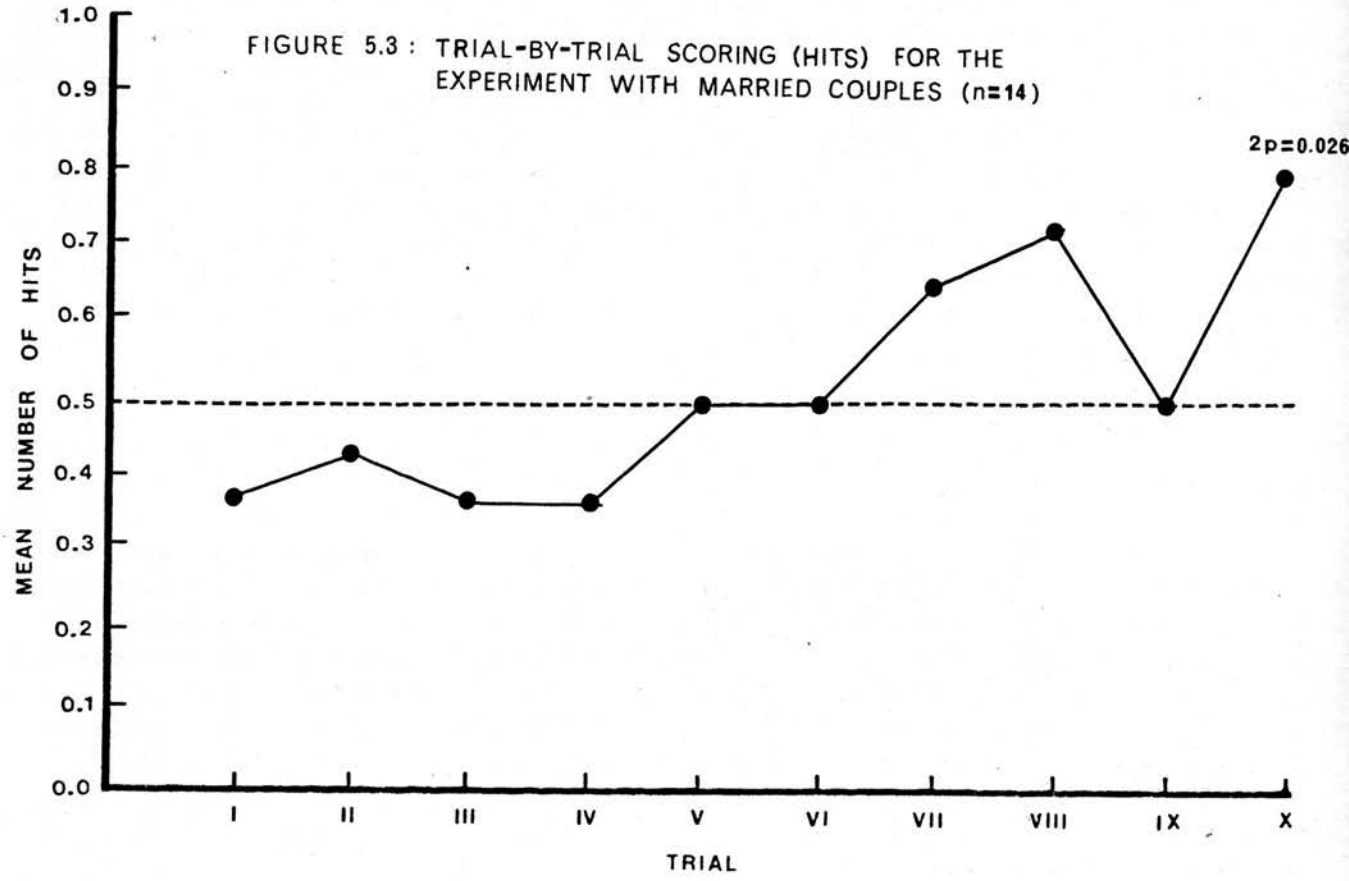
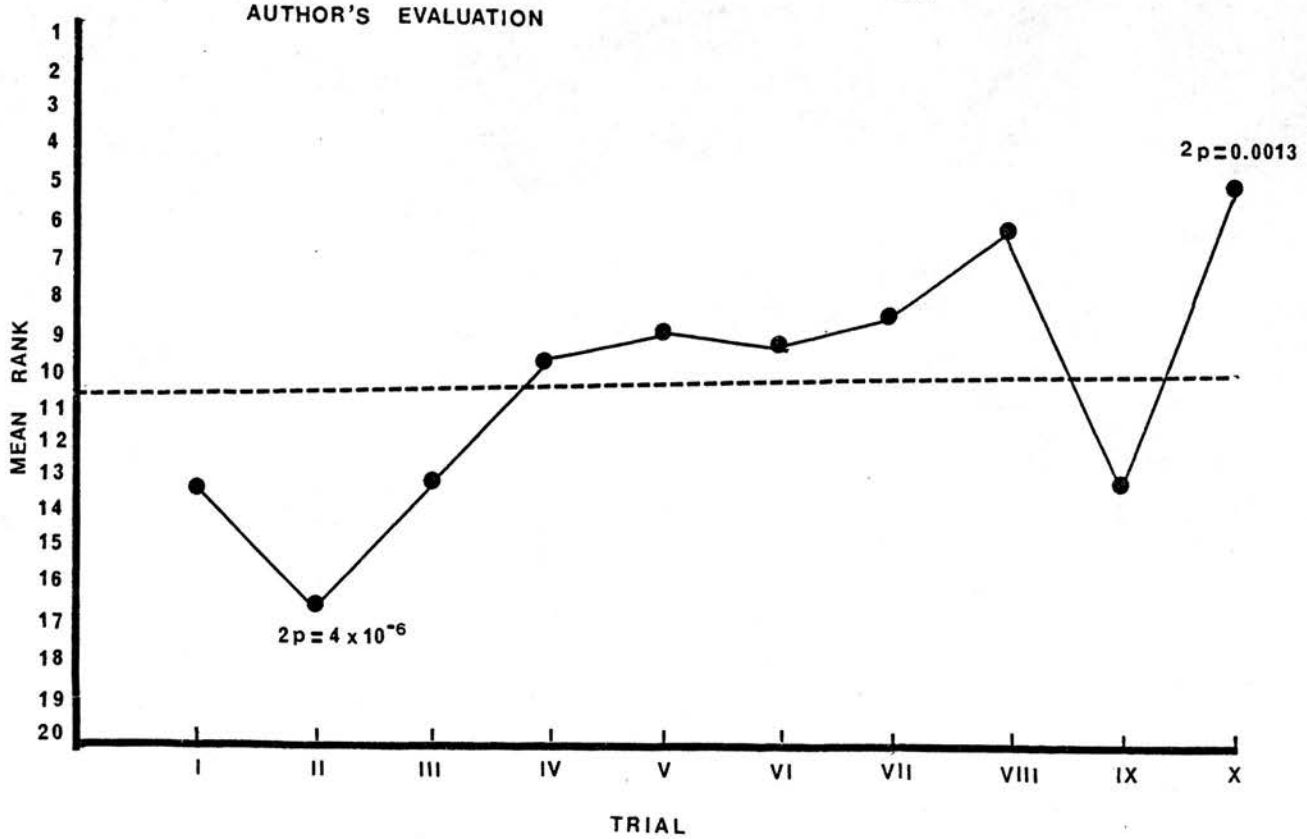


FIGURE 5.3 : TRIAL-BY-TRIAL SCORING (HITS) FOR THE EXPERIMENT WITH MARRIED COUPLES (n=14)



bordering on outright skepticism. Perhaps the one thing that may be said in its favour is that a subsequent experiment (namely, Thalbourne 1980) with a comparably small sample-size, likewise revealed some (equivocal) evidence of a linear trend, together with a good deal of resemblance regarding the structure of the scoring profile (see Figure 5.4), in particular, a tendency for scores to be higher towards the end, but even to the extent of having the 'dip' at the ninth trial: a rank-order correlation between the two scoring profiles portrayed in Figure 5.2 and 5.4 yields a rho of +0.90 (N = 10, 2p = 0.0004).

FIGURE 5.4: TRIAL-BY-TRIAL SCORING FOR THE FIRST AUSTRIA-ICELAND EXPERIMENT (n=7), AUTHOR'S EVALUATION



Thus viewed in the wider context, it might not be altogether wise to dismiss the significant linear trend found here as being no more than a statistical artifact.

(v) Pooling of the Married Couple Data with the earlier Closeness Study.

The final issue that was investigated concerned the question of whether the married couples scored any differently to pairs of subjects who merely shared a close relationship, or who were entirely unacquainted with each other. The married couple data were thus pooled with the data from the original study on emotional closeness, and the three group means were compared using one-way analysis of variance for randomized groups. The results are displayed below in Table 5.3

Table 5.3

Comparison of Mean ESP-Scores for Married Couples, Close-Relaters and Non-Close-Relaters

	SCORE-TYPE	
SUBJECT-GROUP	Mean Rank	Hits
Married Couples	5.609	5.14
Close-Relaters	5.582	5.00
Non-Close-Relaters	5.644	4.23
F, df = 2, 42	0.032	1.678
Probability	0.97	0.20

The results for the mean rank variable could scarcely be more null: the most negative mean is for the non-close group, the most positive for the close-relater group, and married couples scored midway between the two, but the F-test tells us that for all intents and purposes the means may be considered identical. The results for the hits variable are likewise non-significant, but at least the means are in line with prediction that the more 'intense' the relationship, the higher the ESP-score: married couples obtained the highest mean number of hits out of ten, followed by the close-relaters, the rear being brought up by the non-close-relaters. Indeed, an a priori test of the difference between non-close-relaters and the group made up by combining married couples with close-relaters (the mean for this resultant 'undifferentiated close' group being 5.06 hits), reveals a

just-significant t-ratio of 1.825, which, with 42 df, has a one-tailed probability of 0.038. This might have suggested that there could be something after all in the closeness hypothesis, but for the fact that in the Second Sheep-Goat Experiment (Chapter 4), a group of perfect strangers yielded a mean hit-score (namely, 5.29) that was even higher than that for the combined close group.

CONCLUSION

There seems no obvious reason why the present experiment should not have yielded some evidence of psi-hitting in the group of married couples, and thus the most generous conclusion which can be drawn is that even if spontaneous telepathy does frequently occur in every-day life within marital relationships, the phenomenon is not readily imported into the experimental laboratory.

CHAPTER 6

Target-Ranking versus Response-Ranking:

A Methodological Comparison

CHAPTER 6

"The beauty of creation lies in its diversity,
and in the ways in which diverse elements
combine to form meaning and truth".

The Author*

Thus far in the three drawing-reproduction experiments that have been reported, the transformation of degree of target-response resemblance into a numerical score has been performed using one particular method of evaluation: a blind judge is presented with an 'evaluation-group' consisting of all the ten responses drawn by a given percipient, and is asked to rank these drawings for degree of correspondence to just one of the ten target-drawings. Each response-drawing is thus awarded a rank-score representing its degree of similarity to the given target relative to the other nine response-drawings in the evaluation-group.

But an equally valid method of evaluating target-response resemblance is to present the judge with an evaluation-group comprising ten targets, and get him or her to rank these for similarity to just one of the response-drawings. This is the converse of the procedure described above. As a matter of historical fact, Charles Stuart, in all his drawing-reproduction work, habitually carried out both methods of matching. One method he would refer to as "the first matching", the other, "the second matching": in his first report (Stuart, 1942), the former term referred to the case where N responses were ranked against one target, while in the "second matching", the N targets were ranked against one response. (Curiously,

* Inspired by the script of "Is there in truth no beauty?", an episode of "Star Trek" written by Jean Lisette Aroeste.

and confusingly, Stuart subsequently (1945a,b; 1946, 1947) reversed the referents of these terms!.) The present author is not in favour of Stuart's terminology: firstly, it is not at all clear from the names alone which of the two types of ranking is being referred to; and secondly, there seems to be no 'natural' basis for calling one type "the first matching" and the other "the second": he offers no grounds for supposing one type to be prior or superior to the other; both are equally valid. For these reasons, the author proposes that a much more natural and self-evident nomenclature is to call any method in which a set of responses are ranked against a given target a 'response-ranking method', and any method in which a set of targets are to be ranked against one response at a time, a 'target-ranking method'. It is clear, then, that the judging procedure used in all the experiments reported so far in this dissertation, falls under the category of 'response-ranking method'. In Thalbourne (1980) are described two experiments which made use of both methods, but in such a manner as to overlap incompletely. In this present chapter will be reported the outcome of a methodological experiment, in which a set of data (namely, Thalbourne, 1976: the Closeness Study*) which had initially been analyzed using the response-ranking procedure, were re-evaluated according to the target-ranking method. Very generally, the purpose of this comparative study was to examine the extent to which the two methods yield similar results,

---

* together with the data of the five agent-percipient pairs each of whom had scored significantly in the First Sheep-Goat Experiment (c.f. Chapter 3, p. 109.)

both on the level of the actual scores obtained, and the conclusions reached in the process of hypothesis-testing. If the outcomes were very similar, then this would lend confidence to the conclusions that had been previously drawn on the basis of the response-ranking method alone, as well as confirming that the scores derived from each technique are relatively reliable and 'objective' reflections of target-response resemblance. If, on the other hand, there were considerable discrepancy between the two outcomes, this, too, would be valuable information, providing a strong warning that the evaluative procedure entailed a good deal of unreliability.

In addition, however, one formal hypothesis was advanced concerning the results of a target-ranking re-analysis: it was predicted that the degree of inter-judge concordance would be significantly greater in the target-ranking method than in the response-ranking technique. It will be recalled that the evaluation protocol, as described in Chapter 2, requires that the data be ranked by multiple independent judges: that is to say, each evaluation-group is ranked against its master stimulus a total of three times, once by each of three judges, and the resulting rank-scores are averaged to obtain average ranks. Furthermore, Kendall's W coefficient of concordance (Siegel, 1956, pp.229-238) may be calculated so as to determine the degree to which the three judges gave the same rank-order to the objects in the evaluation-group.* It was hypothesized

---

* Mathematically, W is related to the mean of the Spearman rho's computed between all possible pairs of judges, and is thus a 'multivariate' extension of the non-parametric bivariate

that the mean Kendall's W for the target-ranking procedure would be significantly higher than for the response-ranking method, as determined by a paired-samples t-test (or, more conservatively, a Wilcoxon Matched-Pairs Signed-Ranks test).

What is the reasoning that leads to this prediction? The basic premise is that the evaluation task should be in some sense 'easier' for the target-ranking judge than for the response-ranking judge: the latter adjudicator is typically confronted by a set of responses each of which may be vague and ambiguous, badly drawn, depict more than one object, and show some semantic relationship to previous responses (human beings are not very good at generating random sequences, and are subject both to response biases and to idées fixes). For all these reasons a judge may experience an increased amount of difficulty both in recognising the ideas depicted and in discriminating between them in the process of deciding on which ranks to award them. For example, to the extent that two or more responses resemble one another to a non-chance degree, this renders it less easy to decide which has the greater resemblance to the target; and again, in the case of 'multiple responses', one judge may choose on the basis of one part of the response, a different judge on another. An evaluation-group consisting of target-drawings, on the other hand, should present a somewhat easier

---

correlation. It ranges from zero to unity (that is, no concordance at all, to perfect inter-judge agreement), and for ten evaluation-objects ranked by three judges, must be 0.627 or greater in order to be considered significant, that is, in order to conclude that the three judges are applying essentially the same standard.

task to the judge: the targets employed in the author's experiments generally depict just one idea, are adequately drawn, and sets of them should display virtually no response bias and thus be reasonably distinct from each other (by virtue of their topics having been randomly selected from a dictionary and carefully screened so that no two within the same set should be too similar). Thus, in the target-ranking case, it should more frequently happen that multiple judges will agree as to which target goes best with the response, which target next best, and so on.

Now supposing that this hypothesis were correct, several esoteric consequences would follow. When inter-judge concordance (IJC) is low, then the average-ranks will each tend to regress towards MCE, namely, 5.5: for example, if one judge awards a rank of 1, another judge a rank of 10, and yet a third gives a rank of 5, then the average will be 5.33; under such circumstances, it will be rarer that there occur extreme average-ranks (since an average of 1.00 (or 10.00) requires that all three judges give a rank of 1 (or 10), and ex hypothesi this is not happening). Thus, a frequency distribution of the average-ranks would tend to be rather leptokurtic (that is, peaked in the middle, with relatively few extreme scores), with a rather small standard deviation. However, when there is high inter-judge consistency, by definition, the ranks awarded by the three judges tend to agree with each other; in particular, we will observe more agreement about extreme scores, and obtain relatively more average ranks of 1.00 (i.e. three 1's) and of 10.00 (three 10's). This entails that a distribution of such

average-ranks would be more platykurtic (less peaked in the middle, more spread out towards the extremes). Thus, the standard deviation of the distribution will be larger. In short, the higher the IJC, the more variability in the scores.

An increase in variance will have statistical consequences right across the board. One of these is that a given value of a mean rank (that is, the mean of the ten average-ranks awarded to a single agent-percipient pair) will tend to be more significant when it has been derived by means of a response-ranking procedure than by a target-ranking procedure: this is because, with lower IJC under the former method, the standard deviation is correspondingly smaller, and the given mean rank will be relatively further out in the distribution than it would be in a target-ranking distribution. Thus, with particular reference to the Randomization Test for individual scoring, a mean rank which is significant under the response-ranking regime will have to be numerically more extreme from 5.50 in order to achieve a comparable level of significance.

METHOD

A total of 44 judges were recruited for this experiment, most of them being undergraduate Psychology students at the University of Edinburgh. Each was administered the standard judging instructions to be found in Appendix 9, with the minor alteration that they were verbally told, at the conclusion of listening to the tape, that instead of ranking all responses to one target, they would be ranking all targets to one response.

To be re-evaluated in this way were the data of 36 agent-percipient teams, namely, the 31 pairs from the Closeness Study, and the five pairs each of whom had obtained individually significant scores in the First Sheep-Goat Study (see Chapter 3, p. 109). (In the case of these latter five pairs, it was particularly interesting to observe whether the response-ranking significance would hold up under the target-ranking method.)

There was, however, a complication, inasmuch as only 11 target sets had originally been used to test the psi-performance of all 36 subject-pairs (that is, most target-sets had been used for between 2 and 4 different pairs). If the same target-ranking judge were to be presented with a target-set that he has already ranked, even though against a different response-drawing, it is conceivable that his subsequent ranking will be slightly affected by the first, thus violating the Assumption of Independence. For many judges, this problem did not arise, since in the one-hour standardly (and humanely!) allotted for their judging-session they ranked no more than 11 target-sets, viewing each set only once. For the quicker judges, however, who were capable of evaluating more than 11 sets in the hour, an attempt was made to minimize the non-independence problem by: (i) reminding the judge that, since the response presented them on each evaluation-occasion had been produced by an entirely new percipient, the present ranking of the (previously ranked) target-set should be considered completely afresh and independently of the former ranking; and (ii) interposing as large as possible a temporal distance between first and next ranking of the same target-set, which strategy, by all accounts, often had the

effect of destroying the memory of the first ranking. Thus, while the procedure was not absolutely ideal, the author feels that it was satisfactory; the only alternative (namely, having each judge view each target-set once only), while strictly purist, would have had the unpalatable consequence of at least trebling the number of judges required.

RESULTS: PART I

In this first section will be presented (i) the results of testing the formal hypotheses advanced above (see p. 173) concerning the comparative degree of inter-judge concordance observed under the two different ranking-methods, and (ii) correlation coefficients describing the similarity between the ESP-scores as yielded by the target-ranking (TR) method and by the response-ranking (RR) method. In Part II, the data derived from the TR re-evaluation of the Closeness Study will be re-analyzed with a view to testing the process-oriented hypotheses originally examined using RR data (Thalbourne, 1976), and the overlap between the conclusions drawn on the two occasions will then be delineated.

(i) Each of the 36 subject-pairs, as a result of the original RR analysis, had associated with it its own 10 x 10 target-by-response matrix. Each matrix consisted of ten <u>rows</u> of average-rank scores (ten scores in each row): a single row was obtained by having three independent judges rank-order the percipient's ten responses against one target. (See, for example, the partially

filled matrix in Appendix 10.) Analogously, as a result of the subsequent TR analysis, each of the 36 subject pairs now had associated with it an additional 10 x 10 matrix, each matrix consisting of ten columns of average-rank scores (ten scores in each column): a single column was obtained by having three new independent judges rank-order the percipient's ten targets vis-á-vis one response.

For each of these 72 matrices, ten Kendall's W coefficients of concordance were calculated (one for each of the ten triple-rank-orderings of the set of targets (or responses) against each of the ten responses (or targets)): for example, in the case of the RR matrix of Pair 9, for the three judges who ranked all ten responses against the first target, Kendall's W was 0.91; for another three judges, ranking these same responses against the fourth target, W was only 0.47; the average of the ten W's for Pair 9 (that is, $\overline{W}$), was 0.6547. Similarly, for Pair 9's TR matrix, the W yielded when the ten targets were ranked against the first response was 0.87, against the fourth response, 0.93, and the average was 0.7223 (an average in crease of 0.068 from the RR method). In this manner, then, were two mean scores obtained for each pair, namely $\overline{W}_R$ and $\overline{W}_T$. The prediction was that, for the group of 31 Closeness Study pairs, mean $\overline{W}_T$ would be significantly greater than mean $\overline{W}_R$. This one-tailed hypothesis was tested by means of a paired-samples t-test, with alpha at 5% and 30 df. The prediction was verified: mean $\overline{W}_R$ was 0.619, mean $\overline{W}_T$ was 0.655, t = 3.57, 1p = 0.0006. Lest it be objected that a parametric test is inappropriate, the corresponding non-parametric test (a Wilcoxon Test) was performed: for the 31 pairs, the W_T was higher than

the W_R in 22 cases, Z = 3.086, 1p = 0.001. The null hypothesis may thus be safely rejected, and the conclusion drawn that the TR method yields higher inter-judge agreement than does the RR method. Though there are no objective data directly testing the proposition that this phenomenon is due to the greater difficulty of the RR method, the experimenter (who was also the judging supervisor) gained the impression that on the whole, the TR judges performed the task quicker than do RR judges.

The proposition that degree of IJC should be related to the variability of the average-rank scores, also receives some support. The $\overline{W}$ of each of the 31 pairs was correlated with the standard deviation of the distribution of the 1000 Randomization Test means* (the variability of these means -- which were randomly selected from the target-by-response matrix -- should be a reflection of the variability of the 100 component

---

* It may be well at this point to remind the reader that in order to test the significance of a given subject-pair's mean rank statistic, the method adopted in this thesis has been to employ a computer-run Randomization Test (called 'RANMAT'; see Chapter 2, pp.76-77). In this test, the input-data consist of a 10 x 10 target-by-response matrix of average-ranks, and the program randomly selects ten entries from that matrix, repeating this process 1,000 times (with replacement), and calculating the mean of these ten average-ranks on each occasion. The resultant 1000 Randomization Test 'pseudo mean ranks' will of course form a distribution with a given mean, standard deviation and other characteristics. RANMAT compares the actual mean rank obtained by the subject-pair in the ESP-test, with this distribution of Randomization Test means, so as to determine what proportion of the latter are more extreme than is the former, this proportion being taken to be the probability that the obtained mean rank arose 'by chance'. The computer program also produces information as to certain 'centile points' (or 'percentiles') of this distribution of 1000 Randomization Test means: it prints out the value of the pseudo mean rank at, respectively, the 1st percentile, the 2.5th, 5th, 50th, 95th, 97.5th and at the 99th percentile. These particular centile points were chosen so as to give an indication of the value

average-rank scores): when this s.d. had been obtained from a Randomization Test performed on the response-ranking matrix, the correlation with W was r = +.42 (n = 31, 2p = 0.019); when obtained from the target-ranking matrix, r = +.53, (2p = 0.002); and if it is legitimate to combine these two sets of data, for an N of 62, r = +.48, (2p = 0.00008). Thus, as surmised, there seems to be a genuine, though weak, relationship between the degree of inter-judge concordance and the variability of the average-rank scores.

A final analysis compared, for each of the six 'significance-threshold' centile-points of the distribution of Randomization Test means, the mean values obtained using the RR method and the TR method. Weak support was obtained for the speculation that, as a rule, a given value of mean rank must be more extreme from MCE of 5.50 when obtained using a TR method than when using a RR technique, if it is to attain a comparable level of significance: for five of the six centiles (namely, the 1st, 2.5th, 95th, 97.5th and the 99th), the average value at the TR centile-point was more extreme than the average value at the RR centile-point, significantly so for the 99th centile: mean for RR was 7.02 (deviation from 5.50 being 1.52), for the TR centile, mean was 7.09 (deviation = 1.57), paired t = 2.86, 35 df, 2p = 0.007. The effect, though very weak, is thus observable.

(ii) (a) Correlations for Individual Trial-Scores. Each of the 36 subject-pairs has previously had associated with it

---

of mean rank required in order to be significant (at either end of the distribution) at the conventional levels of significance (namely, 0.01, 0.025 and 0.05). These values will differ slightly from matrix to matrix according to the size of the s.d. of the 100 average-ranks comprising that matrix.

ten average response-ranks, representing the ESP-scores awarded for each of the ten telepathy-test trials. (In addition, each of these scores was transformed into a corresponding hit-score and an absolute-deviation score). Analogously, the TR re-evaluation has yielded ten average target-ranks (likewise transformed into hit-scores and absolute-deviation scores). There were thus 360 pairs of trial-scores for each of the three types of ESP-measure.

The correlation between the 360 average response-ranks and the 360 average target-ranks was $r = +.46$ ($2p < 0.001$). There is thus a moderate, though not enormous, degree of similarity between the average rank-scores yielded by the two methods.

The correlation for the hit-scores is rather lower, at $r = +.26$ ($2p < 0.001$): of the 360 pairs of trial-scores, 63% of them were cases where the hit-score remained unchanged from one ranking-method to the other (i.e. hits remained hits, misses stayed misses), but in 37% of the cases a hit became a miss (or vice versa). This low correlation is attributable in large part to hit-scores whose parent average rank-score was near 5.50 (MCE): for if we consider only hit-scores whose parent average-rank was relatively extreme (that is, greater than 8.00 or less than 3.00), then the correlation rises to $r = +.55$ ($n = 101$, $2p = 3.84 \times 10^{-9}$); this is of course because non-extreme average-ranks (that is, those close to 5.50, in the range 3.33 - 7.67) are more likely to move "across the chance-line" upon re-ranking, thereby converting hits to misses (and vice versa).

The absolute-deviation scores are even less stable across

ranking-methods, the correlation being r = +.16 (2p = 0.002). Again, this low degree of association is understandable, since a change of one unit in the average-rank score (e.g. from a 10.00 to a 9.00) is proportionately only 10% of the range of average-ranks (namely, 1.00 to 10.00), but the absolute-deviation score (which would change from 4.50 to 3.50) suffers a proportionately greater change of 23%, since the range of absolute-deviations (when using three judges) is only 0.17 to 4.50.*

By virtue of the relatively low correlations between individual trial-scores derived from the two ranking-methods, we should expect that statistical analyses using the individual trial as the unit are unlikely to replicate significances across ranking-methods.

b) Correlations for Mean ESP-scores of Subject-Pairs. Each of the 36 subject pairs has also previously been allocated a mean rank representing the average of the ten trial-scores, together with a score representing the total number of hits out of ten, and a mean absolute-deviation score. For the 31 Closeness Study pairs, the two values of mean rank obtained from using the RR and the TR method correlated r = +.77 ($2p = 4.1 \times 10^{-7}$). For the two total-hits scores, r = +.66 ($2p = 5.4 \times 10^{-5}$). There is thus considerable similarity between the two values of these ESP-scores as derived by the two ranking methods; we would thus expect that analyses which use as the basic unit a pair's mean rank or their total-hits score will be fairly likely to replicate significances across ranking-methods. The mean absolute-deviation score, however, shows negligible similarity between the RR value and the TR value: r = +.12 (2p = 0.529);

---

* When using three independent judges the range of possible average rank-scores is 1.00 to 10.00, at intervals of 0.33 units. It is thus not possible to obtain an average rank-score of 5.50 (which would require a sum-of-ranks of 16.5 - clearly impossible on an integral ranking system, which does not permit ties); the nearest possible rank-scores to 5.50 are therefore 5.33 and 5.67 (sums-of-ranks of 16 and 17 respectively), and thus the smallest possible absolute deviation score (from 5.50) is of course 0.17.

thus, in this particular case, significances appearing as a result of a TR method are unlikely to appear in a RR technique, and vice versa.

RESULTS: PART II

(i) Subject-Pair Scoring. The individual average target-ranks obtained by each of the agent-percipient pairs on each of the ten telepathy-test trials can be found in Appendix 17 (for the 18 pairs of close-relaters) and in Appendix 18 (for the 13 pairs of non-close-relaters, and (for interest) the five pairs re-evaluated from the First Sheep-Goat Study). Means and other summary statistics for each pair are to be found in Tables 6.1, 6.2 and 6.3. As usual, the mean ranks are arranged in ascending magnitude. It can be seen that for the 31 Closeness Study pairs (Tables 6.1 and 6.2 only), a total of four are significant: of the close-relaters, Pair 11 has a significantly negative mean rank ($\bar{X}$ = 7.47, p = 0.004), a finding which gains strength from the fact the pair obtained a solitary hit and fully nine misses (one-tailed binomial p = 0.011); the same pair also manages to pull off a hat trick by obtaining a significantly high mean absolute-deviation ($\bar{X}$ = 2.87, MCE = 1.90, p = 0.008). Pair 9 also has a significantly negative mean rank ($\bar{X}$ = 7.00, p = 0.024), while pair 12 has a significantly high mean absolute-deviation ($\bar{X}$ = 2.63, p = 0.017). Of the non-close-relaters, Pair 8 had a significantly positive mean rank ($\bar{X}$ = 3.70, p = 0.006). Even considering just the mean rank score alone, the binomial probability of obtaining three or more pairs significant out of a total of 31 at the (average)

Table 6.1 Mean Ranks, Hit-Scores and Mean Absolute-Deviations for Close-Relaters (n = 18)

Pair	Average Rank			Number of hits	Absolute Deviation		
	Mean	S.D.	1p(Mean)		Mean	S.D.	1p(Mean)
30	4.30	2.01	0.080	8	1.80	1.49	0.830
10	4.43	2.68	0.078	7	2.70	1.01	0.051
31	4.57	2.07	0.097	5	1.83	1.33	0.685
14	4.67	2.30	0.154	8	2.10	1.26	0.382
13	4.87	2.10	0.173	5	1.87	1.15	0.493
4	4.97	1.82	0.236	6	1.53	1.11	0.873
5	5.10	1.83	0.284	8	1.47	1.17	0.864
3	5.23	1.98	0.356	5	1.63	1.15	0.692
7	5.43	2.37	0.477	5	1.97	1.32	0.596
16	5.43	2.34	0.464	5	2.03	1.17	0.399
17	5.87	2.37	0.292	5	2.00	1.33	0.393
25	6.00	2.19	0.246	3	2.00	1.01	0.565
22	6.07	1.78	0.205	6	1.30	1.33	0.987
12	6.30	2.80	0.097	3	2.63	1.24	0.017
1	6.33	2.22	0.135	3	2.20	0.89	0.194
20	6.43	2.60	0.081	4	2.40	1.37	0.167
9	7.00	1.79	0.024	3	2.00	1.20	0.613
11	7.47	2.53	0.004	1	2.87	1.43	0.008
Grand Total	5.582	0.91	-	5.000 (s.d.=1.97)	2.019	0.42	-

Table 6.2 Mean Ranks, Hit-Scores and Mean Absolute-Deviations for Non-Close-Relaters (n = 13)

Pair	Average Rank			Number of hits	Absolute Deviation		
	Mean	S.D.	1p(Mean)		Mean	S.D.	1p(Mean)
8	3.70	1.75	0.006	8	2.13	1.32	0.167
27	4.77	2.22	0.176	6	2.20	0.78	0.270
29	5.40	2.48	0.452	6	2.10	1.33	0.310
21	5.50	1.97	0.489	4	1.60	1.15	0.918
15	5.53	2.03	0.483	6	1.70	1.12	0.699
23	5.53	2.24	0.468	6	1.77	1.37	0.678
26	5.53	2.41	0.475	4	2.10	1.18	0.366
19	5.87	2.05	0.323	5	1.87	0.92	0.597
28	5.93	1.81	0.235	5	1.53	1.06	0.855
2	6.60	1.68	0.075	3	1.73	1.01	0.755
6	6.77	1.58	0.035	3	1.70	1.10	0.615
24	6.83	1.67	0.034	2	1.80	1.16	0.547
18	6.97	2.24	0.027	3	2.43	1.11	0.151
Grand Total	5.764	0.91	-	4.692 (s.d.=1.70)	1.897	0.27	-

Table 6.3 Mean Ranks, Hit-Scores and Mean Absolute-Deviations for Significant Pairs from the First Sheep-Goat Experiment (n = 5)

Pair	Average Rank			Number of hits	Absolute Deviation		
	Mean	S.D.	1p(Mean)		Mean	S.D.	1p(Mean)
56	3.97	2.77	0.017	7	2.90	1.26	0.020
34	4.97	2.14	0.249	4	1.73	1.36	0.893
40	5.17	2.29	0.345	5	2.03	1.11	0.574
36	6.27	2.80	0.136	3	2.73	0.99	0.039
57	6.93	1.69	0.014	2	1.97	1.01	0.406

probability of 0.011, is itself significant at p = 0.005, which suggests that at least some of the subjects were individually displaying psi. It is noteworthy that Pair 9 has thus given individually significant results under both RR and TR evaluations, and that Pairs 8 and 11 had both narrowly missed being significant under the former RR method. In this connection, however, we should note that of the five Sheep-Goat pairs previously significant under the RR method, only two (pairs 56 and 57) retained significance on the TR evaluation, though it must be said that at the very least, all the mean ranks were in the same direction (above or below chance) in the two ranking-methods; in fact, and unexpectedly, Pairs 36 and 56 also scored a significantly high mean absolute-deviation under the TR method (see Table 6.3). Thus, while individual significance under one ranking-method can carry over to the other technique, there is no guarantee of it. The actual p-values are nevertheless correlated, which is to say, unlikely to differ grossly between the two methods.

(ii) Group Scoring. The Closeness Study had originally been carried out with a view to testing the hypothesis that the level of ESP-scoring was related to the degree of emotional closeness between agent and percipient (see Chapter 3, p.86) The differences between the grand mean-scores of close and non-close-relaters had therefore been tested by means of Mann-Whitney U-tests and t-tests for independent samples (with 29 df), the hypothesis being one-tailed, with the close group being expected to score higher on all three types of ESP-score.

Analogous tests were carried out for the TR data. Though the group-differences were in the predicted direction, none

of them was great enough to warrant rejection of the null hypothesis. For the mean rank variable, the grand mean for close-relaters was 5.58, for non-close-relaters, 5.76, t = 0.55, 1p = 0.293. For the hits variable, the close mean was 5.00 hits out of a possible ten, the non-close mean was 4.69, and t = 0.45, 1p = 0.327. (In the RR analysis, this score has yielded a non-significant difference by t-test, but a significant one by Mann-Whitney U-test (1p = 0.045); in the TR analysis, however, even this test gave results far from significance. Again, the RR analysis had yielded a significantly negative mean for the non-close-relaters, but this was not found for the TR analysis.) And finally, for the mean absolute-deviation measure, close-relaters averaged above chance at 2.02, non-close-relaters right at chance with 1.90, and t = 0.90, 1p = 0.187. (In the RR analysis, the corresponding mean for non-close-relaters had been non-significantly higher than that for close-relaters.)

None of these six means differs significantly from its MCE. It seems worthy of remark, however, that the mean rank for close-relaters was exactly the same (namely, 5.582) in both RR and TR evaluations, and that likewise the mean number of hits for this group was exactly 5.00 under both ranking regimes! Apart from the mean absolute-deviation measure, the direction of the group-differences was the same in both studies, and similarly non-significant (by t-test) in all cases. Certainly, on the level of group-scoring, the conclusions drawn as a result of testing process-oriented hypotheses are virtually identical under the two ranking methods. This degree of stability -- which is to say reliability -- is highly gratifying.

FIGURE 6.1: TRIAL-BY-TRIAL SCORING (MEAN RANK) FOR THE CLOSE- AND NON-CLOSE-RELATERS

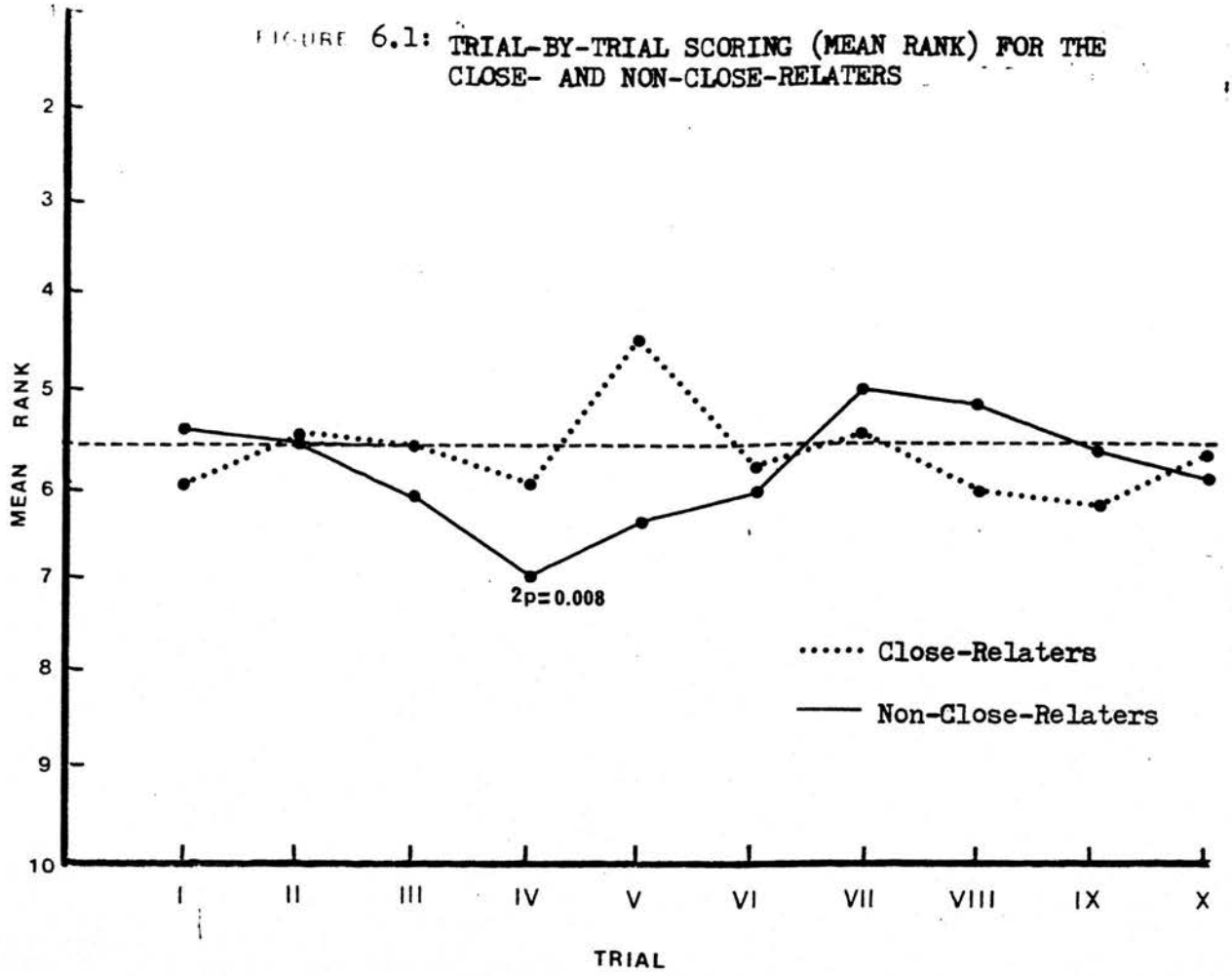
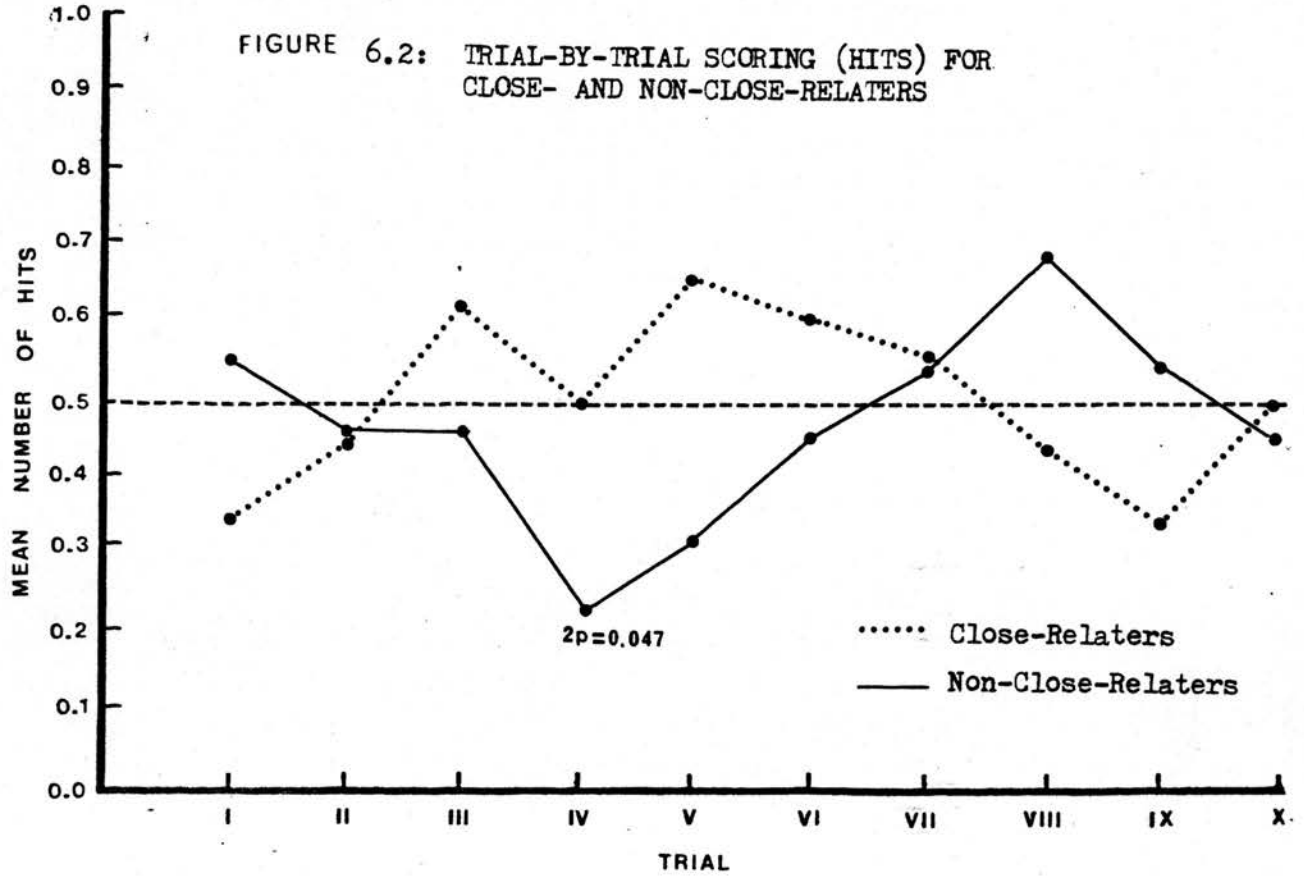
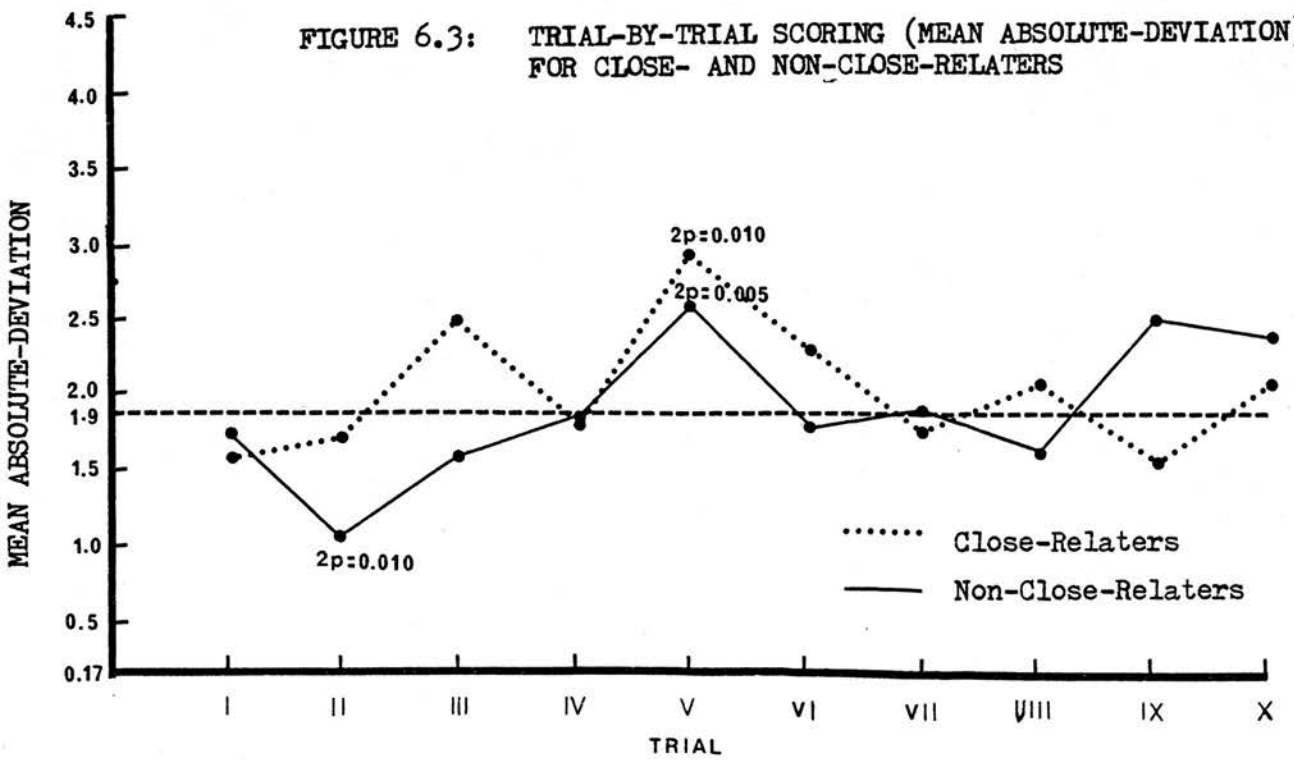


FIGURE 6.2: TRIAL-BY-TRIAL SCORING (HITS) FOR CLOSE- AND NON-CLOSE-RELATERS



(iii) Serial-Position Effect. The mean rank-scores, hit-scores and absolute-deviation scores of close-relaters and of non-close-relaters (separately and combined) were calculated for each of the ten trials in the ESP test. Figures 6.1, 6.2 and 6.3 display these trial-means for each of the two subject-groups considered separately. The data were subjected to repeated measures analysis of variance, in order to examine the possibility that the mean trial-scores changed significantly over time. In contrast to the results obtained using RR data, for neither the average-rank score (Figure 6.1), nor for the hit-score (Figure 6.2) was the between-trials F-ratio significant. However, entirely unexpectedly, the absolute-deviation score (Figure 6.3) did yield a significant position effect for the two groups separately and combined: for the close-relaters F = 2.066, df = 9,153, p = 0.036; likewise for the non-close-relaters,

FIGURE 6.3: TRIAL-BY-TRIAL SCORING (MEAN ABSOLUTE-DEVIATION) FOR CLOSE- AND NON-CLOSE-RELATERS



$F = 2.066$, $df = 9,108$, $p = 0.043$; and for the combined group of 31 pairs (graph not illustrated), $F = 2.580$, $df = 9,270$, $p = 0.007$. What this result seems to imply is that the average-rank scores changed significantly over time regarding the degree to which they deviated from MCE of 5.50. At one stage in the ten-trial test (trial 2, particularly), the average-rank scores were uncommonly close to MCE of 5.50, while at other stages (notably, trial 5), the scores tended to deviate markedly from the expected value of 5.50 (this latter being expected if some subjects psi-hit and others psi-missed). Since this peculiar effect has never before been observed in the author's experiments, one is tempted to put it down to chance, yet this 'chance' effect manifested itself in both groups of subjects!

Trend analysis (using orthogonal polynomial coefficients) yielded a large number of significant results, but the pattern of these results can only be described as perplexing. For the average-rank score, the non-close-relaters yielded a significant cubic trend ($p = 0.022$). For the cognate hit-score, the non-close-relaters' cubic trend was marginally significant ($p = 0.067$), but in the case of the close-relaters it was the quadratic component that attained significance ($p = 0.037$); indeed, as inspection of Figure 6.2 will reveal, this U-shaped trend is inverted -- quite opposite to what was found using RR data! But it is when we come to the absolute-deviation score that trend analysis yielded the most significance. For the 18 close-relater pairs analyzed alone, the quadratic component was significant ($p = 0.022$). For the 13 non-close-relater pairs considered alone, both linear and quintic components were sig-

nificant ($p = 0.017$ and 0.035 respectively). However, the effect of combining the two groups is to yield a trend not observed in either group, namely a significant quartic (W-shaped) component ($p = 0.044$). Thus, while both subject-groups appear to have given evidence for a significant change in variance over the ten-trial test, the actual pattern of these fluctuations is different for each group. It must be emphasised, however, that these effects were totally unpredicted, and that nothing comparable has ever been observed in studies using the response-ranking method. It would seem safest, then, to withold judgement concerning the genuineness of the effect until additional target-ranking data accumulate: it may then be possible to discern patterns recurring in different experiments, just as happened in the case of the U-shaped curve with the RR data.

The failure to replicate the significant U-shaped trend found using the response-ranking method (see Chapter 3, p. 102.) can be attributed in large part to the reasons described above on p. 183: from one ranking-method to the other, individual trial-scores are subject to a very great degree of variation, and thus analyses which employ the trial-score as the basic unit (as does the test for serial-position effect) are unlikely to replicate significance across methods. The level of significance attained in the response-ranking data was, in any case, not much less than 5% -- a fact which suggests that the effect was rather weak. (This suggestion tended to be borne out in subsequent (response-ranking) studies, where the quadratic component in trend analysis was significant only when the sample-size was fairly large.) So this fact too -- the sheer weakness of the U-shaped effect itself -- may help

to account for the lack of significance when using the TR method.

On the other hand, this explanation does little to assuage the sense of uneasiness one feels at the thought that there is so little resemblance between the trial-by-trial scoring profiles yielded by the RR and by the TR methods: using two supposedly 'equivalent' instruments to analyze the same set of data, totally different scoring patterns emerged: we cannot even predict with any certainty that a trial-mean whose value lies above chance on one method will likewise be above-chance on the other method! So how do we know what is the 'real amount' of similarity between target and response, if the two methods yield inconsistent values?

Perhaps a clue to resolving this dilemma lies in understanding the very meaning of the terms 'target-rank' and 'response-rank': a response-rank value of 1.00 indicates that three independent judges have decided that the given response-drawing is more like the master-stimulus target than are any of the other nine responses; a target-rank value of 1.00, on the other hand, means that, in the opinion of three independent judges, the given target is more like the master-stimulus response than are any of the other nine targets. Thus, for the two methods, the evaluation-group of comparison stimuli (compared with which the given response (or target) is adjudged to have greater resemblance to the master) is quite different: in the RR case it consists of the subject's nine other response-drawings, while in the TR case it comprises the nine other targets. There is thus no guarantee that the rank-score awarded to a given response when ranked amongst the other nine responses, will be the same as the rank-score awarded to that response's own

target when that target is ranked amongst the other nine targets. Indeed, perfect coincidence of rank-scores will generally be achieved only if the degree of target-response resemblance is 'in fact' very large,* and this will occur only if paranormal cognition is taking place very strongly on each individual trial. (For a case of perfect correlation between target-scores and response-scores, see the full paper of Delin's (1977) re-analysis of Uri Geller's picture-guessing data (Targ & Puthoff, 1974a).) Until that situation actually occurs in our experiments, individual trial-scores will suffer from a certain amount of 'quantum indeterminacy' ; fortunately, these random fluctuations at the 'subatomic' level do tend to cancel each other out at the higher, 'macroscopic' levels of analysis: the mean of a subject-pair's ten trial-scores tends to be fairly reliable across the two ranking methods, and the grand mean of a group's mean is even more stable.

(iv) The Correlation Analysis, and the Sheep-Goat Effect.

In the introductory section of Chapter 3, a detailed description was provided concerning a rather large analysis, the aim of which was to investigate whether the ESP-scores obtained in the Closeness Study were correlated with any of a large number of psychological variables. The primary conclusion drawn from this analysis was that the data provided strong evidence of a sheep-goat effect, at least in those dyads where agent and percipient were not involved with each other in a

* Strictly speaking, 'very large' should read 'very deviant from chance degree of resemblance, either above it or below'.

mutually close relationship. The evidence further suggested that the sheep-goat status of the percipient was more determinative of ESP-score than was the status of the agent; (subsequent experimentation, however, caused us to question this secondary conclusion, in light of the fact that the sheep-goat effect proved more reliable in the agents than in the percipients.) Be that as it may, these two inferences were made on the basis of a response-ranking analysis of the data; it was therefore of very great interest to learn whether the same conclusions would be drawn when the target-response resemblances were evaluated using the TR method.

In the Results section above (Part I, (ii)(b)), it was observed that from one ranking-method to the other there was considerable similarity between the two obtained values of the ESP-scores mean rank and total-number-of-hits ($r = +.77$ and $+.66$ respectively). So even though between 41% and 56% of the variance in each of these scores is unaccounted for, it was nevertheless to be expected that when correlation coefficients were computed between the psychological variables and the new, TR ESP-scores, they would tend to be reasonably similar to those obtained using RR ESP-scores. Very gratifyingly, this is precisely what happened. In a manner entirely analogous to the original correlational analysis, a grand total of 1170 coefficients were computed. On the RR analysis, 145 of these values -- or 12.4% -- had been significant at the (two-tailed) 5% level or better (compared to a mere 3% significant correlations in the best of three Shuffle Control tests (cf. Chapter 4)).

Table 6.4 Correlations between Mean Rank and the Sheep-Goat Scale and Component Items, Closeness Study (Target-Ranking Data)

	AGENTS			PERCIPIENTS		
	All A (n=31)	CR A (n=18)	NCR A (n=13)	All P (n=31)	CR P (n=18)	NCR P (n=13)
Sheep-Goat Scale 3	-.60††	-.48*	-.84††	-.59***	-.59**	-.63*
Belief in ESP	-.05	-.25	+.40	-.31	-.30	-.43
Personal experience of ESP	-.29	-.28	-.27	-.38*	-.37	-.31
Believe Self to be Psychic	-.16	-.08	-.28	-.50**	-.46	-.63*
Have had a Hunch	-.47**	-.36	-.74**	-.42*	-.40	-.42
" " Premonition	-.33	-.26	-.58*	-.42*	-.42	-.43
" " Precognitive Dream	-.54**	-.48*	-.61*	-.53**	-.44	-.75**
" " Vision	-.33	-.44	_.31	-.26	-.29	-
Believe in life after death	-.17	-.03	-.30	+.09	+.12	-.03
Believe in Spirit Contact	-.35	-.37	-.14	-.42*	-.26	-.47
Telepathic experience with Partner	-.27	-.37	-	-.32	-.50*	-
Telepathic experience with Non-Partner	-.33	-.15	-.63*	-.59†	-.75††	-.31
Have had a Telepathic Experience	-.36*	-.23	-.63*	-.55***	-.66**	-.31
Estimated Probability of Success	-.43*	-.36	-.36	-.45*	-.38	-.54
Belief that ESP occurred	-.11	-.06	-.35	-.13	-.10	-.16
" " was inhibiting factor	+.06	-.08	+.32	-.20	-.26	-.09

Note:
* : $2p \leq 0.05$
** : $2p \leq 0.01$
*** : $2p \leq 0.001$
† : $2p = 0.0004$
†† : $2p = 0.0003$

'A' = 'Agents'
'P' = 'Percipients'
'CR' = 'Close-relater'
'NCR' = 'Non-Close-Relater'
_ = correlation non-computable due to lack of variance.

On the TR analysis, the number of significant correlations yielded was virtually identical to that of the RR analysis -- namely, 140, or 12.0%. Indeed, 62 of the 140 variable-pairs that were significant on the TR method, had also been so on the RR analysis. On the gross level of 'headcounts', then, the degree of overlap between the two ranking-methods is thus impressive.

Just as we found in the RR analysis, the ESP-score called 'mean absolute-deviation' accounted for a disproportionately small percentage of the significant correlations: 89% of the significant coefficients were between various psychological variables and either mean rank or hits. Again, as far as the total number of correlations is concerned, sheep-goat variables accounted for just 54% of the 1170 coefficients computed; and yet, in the TR analysis they accounted for 81% of all the significant correlations (as compared with a figure of 86% in the RR analysis). We may therefore tentatively conclude that the TR analysis has yielded evidence of a sheep-goat effect. Thus, on this very general level, the TR results mesh beautifully with those obtained using RR data.

Table 6.4 displays in detail the correlations between the mean rank variable and the various sheep-goat variables (negative correlations indicating positive ESP-scoring on the part of the sheep, negative scoring by the goats).* It can quite readily be seen that, very strikingly, the correlations

---

* The correlations between mean rank and the Sheep-Goat Scale-scores are all Pearson r's; between mean rank and the Sheep-Goat items, Spearman rho's.

are almost all of the same sign: of the 93 coefficients, only five are positive (indicating a 'goat-sheep effect'), and none of these is significant; of the 87 negative correlations, favouring the sheep over the goats, fully 30 are significant, the highest of these being the correlation between mean rank and the Sheep-Goat Scale-scores of the 13 non-close-relater agents ($r = -.84$, $2p = 0.0003$).

At this point the reader might like to review Table 4.2 -- the corresponding correlations for the response-ranking data. Careful comparison of Tables 4.2 and 6.4 will reveal that while both yield striking evidence of a sheep-goat effect, they do nonetheless differ in important ways regarding the distribution of this effect over the six subject-groupings. Using RR data, it had been concluded that the sheep-goat status of the percipient was rather more strongly correlated with ESP-score than was that of the agent, and that within the percipients themselves, their correlations tended to be significant only if they were non-close-relaters. The target-ranking analysis, however, gives a rather different story: though the emphasis is not reversed, there is nevertheless a shift in emphasis away from the non-close-relater percipients. For the group of 31 percipients as a whole (Table 6.4, column 4), the significant correlations are comparable in quality and only slightly superior in number to those of the 31 agents (column 1). Again, the process of dividing the percipients into the 18 close-relaters and the 13 non-close-relaters, yields results that do not much favour one group over the other. When, on the other hand, we separate the 31 agents into the 18 close-relaters and the 13

non-close-relaters, there is a slightly more pronounced tendency for the non-close-relaters to carry the day in number and size of significant correlations. But perhaps the most interesting point is that had we been asked to conjecture, on the basis of the Sheep-Goat Scale/ESP correlations, which of the six subject-groupings seemed most likely to replicate in a future study, the choice would in all probability have fallen upon the non-close-relater agents ($r = -.84$), and certainly not on the non-close-relater percipients (the choice made in the RR analysis). In retrospect this is highly ironical, in view of the fact (let us remind the reader) that the conclusion reached after two follow-up studies of the sheep-goat effect (Chapters 3 and 4) using a response-ranking technique was that: (i) a necessary pre-condition for the manifestation of the sheep-goat effect in a GESP test was that agent and percipient not possess a close emotional relationship with each other, and that (ii) the effect appears more reliably if we predict ESP-scores on the basis of agent's Sheep-Goat Scale-score rather than that of the percipient. The irony, therefore, is that the TR analysis of the original data has suggested precisely the same conclusion to which we were ultimately led using RR data: namely, that the best candidate for showing the sheep-goat effect is the non-close agent!

This inference tends to be confirmed as a result of carrying out on the data the technique known as multiple correlation. It may be recalled from Chapter 4 that in a more formal, more elegant attempt to demonstrate that there was a genuine relationship between the ESP-scores and the psychological predictors

(and not just pseudo-significances generated by 'data-scrounging'), the 33 psychological variables (see Table 3.1) were reduced by factor analysis to just nine relatively independent factor-scores (see Table 4.1); these latter were then used, as a set, to predict each of the three different types of ESP-score for each of the six subject-groupings. In the response-ranking data, only one of the resultant 18 multiple correlations had been significant, namely, when the dependent variable was the mean rank scores obtained by the 13 non-close-relater percipients.

Analogously, 18 such analyses were carried out using the TR data, and four of these were significant (but, tellingly, none of these four involved the non-close-relater percipients!). The most significant of these four multiple correlations was the one describing the degree of linear association between the nine factor-scores and the mean rank scores obtained by the 13 non-close-relater agents: the multiple r was 0.993, $F = 24.20$, $df = 9,3$, $p = 0.012$. Additional F-tests with 1 and 3 df indicated that the beta weights were significant for four of the nine factor-scores: sheer number of psychic experiences ($F = 146.91$, $p = 0.0012$), reported experience of telepathy ($F = 56.90$, $p = 0.005$), anxiety during the ESP test ($F = 12.02$, $p = 0.04$), and extraversion ($F = 11.71$, $p = 0.042$).*

---

* For the sake of completeness, it should be added that the other multiple correlations which were significant were, in order of significance, (i) the prediction of the mean absolute-deviation score of the non-close-relater agents ($p = 0.021$); (ii) prediction of the mean rank score of the 31 percipients as a group ($p = 0.031$); and (iii) prediction of the mean rank score of the 18 close-relater percipients ($p = 0.032$). One would therefore want to temper the conclusion that the sheep-goat effect occurred only in the non-close-relater agents by qualifying it with the statement that they seemed to show the best, most significant evidence for the effect.

DISCUSSION

What, then, are the most important conclusions that may be drawn from this comparative methodological study? The first is that compared with the response-ranking technique, the use of the target-ranking technique with multiple judges results in significantly higher inter-judge concordance. Though the absolute magnitude of the difference in concordance from one method to the other is not overly large in real terms -- the grand mean W increasing from 0.62 to 0.66 in the present study -- nonetheless the importance of this new finding cannot be overstated. We argued in the Methodology chapter (Chapter 2) that the use of multiple judges is to be prefered over a single-judge scheme on the grounds that the former led to ESP-scores which were more valid and more reliable -- that is to say, more 'objective' (in the sense of 'consensual') measures of the 'true' degree of target-response resemblance in the data, and as a consequence more stable, in that different sets of judges should award acceptably comparable scores. In the present author's opinion, some of the problems encountered in attempting to replicate ESP effects found in free-response research may well stem from insufficient reliability in the scores derived from the judging process (quite apart from the additional nemesis of the unreliability of the psi phenomenon itself!). Anything which tends to increase the reliability of the judging-scores without at the same time sacrificing their validity is therefore to be heartily recommended. The use of multiple judges will produce just such an effect. So, it now seems, will the use

of a target-ranking method instead of a response-ranking technique: increased inter-judge concordance should mean increased stability of scores. The author has been so influenced in this regard that in any future drawing-reproduction tests he carries out where a choice has to be made between the TR and the RR method, the choice will fall upon the former.

The second major conclusion that we may draw from this experiment is that when trained upon the same set of raw data, the target-ranking and the response-ranking methods yield ESP-scores that are unlikely to be identical and yet which are at least 'reasonably' similar to one another. The similarity is greater for the primary ESP-score known as the 'average-rank' than it is for the secondary ESP-scores that are derived from the value of the average-rank, namely, the hit-score and the absolute-deviation score. The similarity is also generally greater when we are dealing with the mean of a set of ESP-scores than when we are treating each individual trial's ESP-score as the basic unit of analysis.

It follows logically from this second major conclusion that if the raw ESP-scores yielded by the two ranking-techniques are likely to be similar, then too similar inferences are likely to be drawn when those data are used to test hypotheses about the existence and nature of psi. This is approximately what happened when the raw data of the Closeness Study were evaluated using the two ranking-methods. The principal inferences drawn as a result of a RR analysis of these data were as follows:

(a) there was a very slight suggestion that certain

agent-percipient pairs may have produced individually significant evidence of psi;

(b) close-relaters in general scored non-significantly higher than did non-close-relaters (who apparently psi-missed), thus failing to provide a significant confirmation of the hypothesis that ESP-scoring is related to the degree of emotional closeness between sender and receiver;

(c) there was a significant change in level of group-scoring over the ten trials in the ESP-test, this change conforming significantly to a U-shaped curve, such that the group tended to score above chance at the beginning and end of the test but below chance in the middle;

and (d) there was strong evidence of a sheep-goat effect, particularly amongst the non-close-relaters who were percipients.

The corresponding inferences based on the TR data were, seriatim:

(a') there was good evidence that certain subject-pairs gave significantly extra-chance scores;

(b') the 18 pairs of non-close-relaters scored higher than did the 13 pairs of non-close-relaters, but in no case did the group-difference approach significance; neither did any of the group-means independently deviate from chance to a significant degree; the Closeness hypothesis was therefore not supported;

(c') there was some sort of change in group-scoring over the ten trials, such that at one stage the members of the groups obtained trial-scores remarkably close

to chance, while at other times their scores were markedly deviant from chance (some subjects scoring very high, others very low);

and (d') there was excellent evidence of a sheep-goat effect, slightly stronger in the case of the non-close-relater agents.

Comparison of these two sets of conclusions, point for point, will show if not an identity, at least a quite reasonable degree of dove-tailing, particularly for point-pairs aa', bb' and dd'. The most marked discrepancy is between conclusions c and c', concerning the nature (though not the existence) of some sort of serial-position effect; possible causes for this discrepancy have been discussed in the Results section; yet even in this case, on a very general level, one may note that the same moral can be drawn from both the TR and RR analyses, namely, that changes of some sort may occur in scoring during the course of a multi-trial test, and that these changes are frequently obscured by examining only the average level of performance over all trials.

On the very general issue, then, of whether ESP was occurring at all in the Closeness Study, the two ranking techniques have yielded complementary rather than conflicting results: the verdict favours the paranormal hypothesis. On the more specific level, however, (namely, the question of just how and when and in whom ESP was manifesting), the results are rather more divergent. But at least this methodological study has given us some guidance as to what sort of conclusions are reliable and which are less so: to wit, analyses based on trial-scores

are the least likely to be confirmed from one ranking-method to the other; while analyses based on higher-order scores (subject-means, group-means) are rather more likely to yield stable inferences. If, in other empirical studies, we have only one ranking-method with which to evaluate data, we now have some idea as to what degree of confidence to place in the various types of significance-analyses we undertake.

A word of caution is in order at this point. The results obtained from examining the stability of ESP-scores and analyses from one ranking-method to the other, may well be dependent upon such factors as the size of the evaluation-group used in the judging process, and the number of independent judges used to rank-order the objects in that group. For example, it may be that if the evaluation-group was smaller than 10, the RR procedure would be easier, and yield a degree of inter-judge concordance non-significantly different from that observed under TR conditions. Again, the use of more than three judges may well yield a better level of concordance than the typical 0.64 that we tend to observe in this series of experiments, and again this fact may obliterate the differences we found between the two ranking methods. It would be a most interesting project to manipulate systematically either or both of these variables with a view to observing their effects on the outcome of a methodological comparison: until such time as this is done, we might well be advised to confine the generality of our conclusions to experiments that use a trinity of judges, each of whom is to rank-order an evaluation-group containing ten objects.

Another possible variable in this regard is the degree to which ESP is itself operating in the experiment under consideration. We mentioned above that the more strongly psi was occurring in the data, (i.e., the more extra-chance the degree of target-response resemblance), the more likely that the two methods will yield closely similar results. But does the converse also hold? Is there no overlap between the results -- no correlation at all -- if there is no evidence for ESP? Surely not! As point-pair bb' above showed, the two ranking-techniques can concur on the matter of non-significance as well as significance. This is as it should be: the fact that, when ten targets (or responses) are ranked against one response (or target), three independent judges agree significantly in the ordering they make, implies only that they are all applying essentially the same standard to the data (that criterion presumably being, in line with their instructions, the imperative "rank-order on the basis of topographical and/or semantic similarity"). But this is quite a different issue to the question of whether the correct response (or target) is ranked highly against its own corresponding target (or response). Likewise, the fact that individual average response-ranks correlate +.46 with average target-ranks (and the fact that this correlation is astronomically significant for an N of 360), is not itself evidence of an ESP effect, but implies only that the response-ranking judges awarded a 'rating' of target-response resemblance which was reasonably 'objective', and such that a trio of TR judges were able to detect, as evidenced by their giving approximately comparable ratings. Speaking extremely loosely, it

is as if there existed a real 'quantity' of target-response similarity, and the significant correlation indicates merely the useful information that two different measuring instruments were able to yield the same estimate of that quantity. The objectivity of the score is one thing; its evidentiality, quite another.

If the author were asked which of the ranking-methods is to be preferred, the answer would have to be "the TR method": at least in the present experiment, this technique has led to detectably greater inter-judge reliability -- always a good commodity to have in abundance; and on the pragmatic level, it seems easier and less time-consuming for judges to carry out. But before closing this chapter, we should perhaps address the normative issue of whether both ranking-analyses ought to be required of any adequate assessment of a set of free-response data. At the current stage of research, one can only say that in general, each ranking-method leads to approximately similar conclusions, and that therefore the use of just one of the methods can, with appropriate caveats, be relied upon to give reasonably generalizable results. On the other hand, since free-response evaluation entails so much 'error variance', it may under certain circumstances be wisest to employ the two methods in concert: very weak ESP effects tend not to survive across the two methods, but comparatively strong ones should. So in the end it comes down to whether the experimenter prefers to be a 'Type I errer' or a 'Type II errer': if she or he is prepared to err occasionally by concluding that psi is present when

in reality the images are mirages, then the use of a single method may suffice; if, on the other hand, the researcher has the resources -- in terms of time and man-power -- and is concerned to draw conclusions only on the basis of exceptionally solid evidence, thus giving chance the benefit of the doubt, then the use of the two ranking-methods side-by-side is to be recommended.

CHAPTER 7

Discussion, and Conclusions

CHAPTER 7

"When infant mumbles,
and child fumbles,
the body yet
is a marionette:
'tis the newborn soul
gaining control!"

The author

In each of the four preceding chapters has been reported a discrete empirical study of the paranormal cognition of drawings. The time has now come for us to undertake a more global overview of these experiments, in an attempt to summarize the light which they may collectively throw on our understanding of the properties of ESP. After much panning, we seem to have discovered a little gold dust -- perhaps even a nugget -- glittering at us from all the dross and dirt; but it will be necessary to examine the finds closely so as to be sure that they are genuine, and not fool's gold! The discussion will therefore include, where appropriate, the issue of the adequacy of the experimental design, in an attempt to see whether the putative ESP effects found are explicable in terms of non-psi modes of influence. Provided that we can justify the claim that ESP did take place in these experiments, we may then go on to discuss the implications of it in the wider context. The exposition will proceed under three major headings, namely, (i) the Closeness Hypothesis, (ii) Serial-Position Effects, and (iii) the Sheep-Goat Effect.

(i) The Closeness Hypothesis. When the author first embarked upon parapsychological research, he decided to test the validity of the popularly-held belief that emotional closeness between

agent and percipient in some way enhances the occurrence of ESP (particularly psi-hitting). A ten-trial picture-guessing experiment was therefore carried out in which the GESP-scoring of 18 pairs of "close-relaters" was compared with that of 13 pairs of "non-close-relaters". The results of this study (initially reported as Thalbourne, 1976) failed to demonstrate that the close pairs scored higher than the non-close dyads in any convincingly significant fashion. (If anything, the non-close pairs yielded more extra-chance scoring, albeit in the negative direction (see the introductory section of Chapter 3, in this thesis)). In the study described in Chapter 6, the resemblances between the targets used and the responses produced in this experiment were re-evaluated by means of a target-ranking (instead of a response-ranking) procedure; the results of this parallel analysis only served to confirm the prior conclusion that emotional closeness between agent and percipient is not sufficient to ensure that level of ESP-scoring will be enhanced.

On the not unreasonable supposition that perhaps the marital relationship would, of all close relationships, be one of the more likely to be conducive to ESP, a test was carried out using as subjects 14 couples each of whom had been married less than ten years (Chapter 5). Alas, the yield of evidence for psi was so low in that study that it has come to hold the dubious distinction of being the author's least successful experiment ever conducted! Indeed, one of the very few faint glimmers of significance espied therein was the suggestion that, far from there being a tendency to increase in psi-hitting with length of time married, quite the opposite trend: an increase in the tendency to score

negatively! If this finding ever proved replicable it would have disastrous implications for the future of Matrimony! Though one could try the reader's patience by enumerating a lengthy list of reasons why close dyads should have performed so poorly in these studies (e.g. the absence of spontaneity -- induced by the awareness of being tested; the triviality of the target-information compared with real-life crises, and so on), it would seem most logical simply to conclude that popular belief has not in this case provided us with a useful clue to enhancing level of ESP-scoring in laboratory conditions. Closeness of relationship may yet influence the occurrence of spontaneous ESP, outside the laboratory, but once within the walls of the latter it often seems to have negligible effect. Furthermore, emotional closeness is not merely insufficient but unnecessary for the occurrence of GESP: as will become clearer below, other studies reported in this dissertation have demonstrated that barely-acquainted strangers can manifest ESP (the direction of scoring apparently depending on certain other psychological characteristics of the two persons involved). Laymen are very apt to exhibit surprise at these findings, which only goes to show how deeply entrenched is the belief that emotional closeness is an essential pre-requisite for any psi-communication.

(ii) Serial-Position Effect. The test used to examine ESP-performance in all these studies required that each subject-pair have their behaviour 'sampled' on each of ten successive trials -- three minutes each trial -- and without either agent or percipient receiving any feedback (until after the tenth trial) as to the

apparent success of their attempts at "telepathic communication". Because of this repeated measures design (as opposed to a design where each subject-pair contributes just a single trial), and since too the entire test extended over a period of 30 minutes, it seemed plausible to suppose that the level of ESP-performance might not necessarily remain constant over all trials. Changes in level of scoring -- particularly declines from high to low scores -- have so frequently been observed in lengthy sessions of psi-testing that they have almost become a 'fingerprint' of the paranormal process (Beloff, 1981). The averaging of performance over trials may well obscure these internal effects: by pooling high-scoring trials with low, the result may be a 'cancellation' effect, yielding a mean score close to chance.

It was with this thought in mind that, for every experiment, one of the routine planned analyses was a breakdown of the trial-by-trial scoring, with a view to testing for 'serial-position effect' -- an effect in which the level of group-scoring on a particular trial is related to the position that that trial holds in a series. Such an effect may manifest in (a) a significant alternation between high and low-scoring, and/or (b) a significant degree of conformance between the ten trial-means and a meaningful mathematical function (linear, quadratic, or an even higher-order polynomial).

The search for such effects did indeed yield some significant results. Of the four (response-ranking) studies reported in this thesis, only one gave a significant <u>change</u> in level of scoring over trials, but no less than <u>three</u> gave significant evidence of a mathematical <u>trend</u>: of these three, the one with the smallest

sample-size yielded a linear trend (the Married Couple Study: n = 14), while the other two (the Closeness Study and the Second Sheep-Goat Experiment), with n's of 31 and 27 respectively, both yielded a U-shaped curve. Imagine the author's delight when, as a result of re-analysing a ten-trial drawing-experiment conducted by Mittenecker & Schulter (1978) between Austria and Iceland (n = 7), he discovered both a significant change in level of scoring over trials and a significant linear trend (remarkably similar to the trend found in the Married Couple Experiment, which likewise used a rather small sample-size). (See Thalbourne, 1979a.) Unfortunately, the one experiment subsequently conducted which had as its primary hypothesis the occurrence of a serial-position effect, failed to provide any evidence whatsoever for such an effect (Thalbourne, 1980).

So what, then, can we predict? If we can trust so small a sample size as six studies, the 'trend of the trends' that seems to emerge is that the U-shaped curve will not be seen if, amongst other things, the sample-size is not at least 27 subject-pairs, while with any less then this number a trend may appear but be more likely linear. One may recall from Chapter 4 that there was some post hoc evidence that the U-shaped curve observed (at least in the pooled results of three studies in each of which there had been some evidence of this curve) was in fact a result of a linear decline in scoring from first trial to sixth, followed by an upswing to essentially chance-scoring. It would have been very neat, therefore, if the linear trends observed in the Married Couple and Austria-Iceland Experiments had also been declines, but in fact they are inclines, so one is not much closer to understanding the pattern.

Is there any way in which these effects might have been produced by normal rather than paranormal means? As noted in Chapter 2, there were in general two possible avenues by which sensory leakage might have occurred in these experiments: the first is from experimenter to subjects during the pre-test social interaction (experimenter was somewhat aware of what the targets were of); and the second is from experimenter to judge in the course of the evaluation process (experimenter had previously seen the correct target-response matches). Now U-curves are said to be typical of memory-tasks in experimental psychology -- the first and last items of a sequence to be memorized being the most likely to be correctly recalled. It may be, then, that the experimenter tended to remember better what the first and last targets were, and remember only poorly the targets in-between, and therefore differentially emit cues to the subjects prior to their being tested; again, experimenter may have been more likely to recall which response was actually made for the first and for the last targets, and thus have unconsciously influenced the judges to rank highly the appropriate response-drawing in the evaluation-group.

Against this hypothesis, however, must be set the following facts: (i) two of the observed trends are not U-curves but linear inclines, with the first trial giving one of the most <u>negative</u> means; (ii) one has also to explain why psi-missing rather than merely chance-scoring occurred at the bottom of the U-curve; and (iii) perhaps most cogently, neither of these two sensory cueing possibilities existed in the two Austria-Iceland experiments (Thalbourne, 1979a, 1980), and yet in the first of these, excellent

evidence of a serial-position effect was found.

If we may therefore accept the greater likelihood of the psi hypothesis over sensory-cueing artifact, then a few thoughts may be in order as regards the causation of this effect. Several confounding variables may be identified, any or all of which may be implicated in the changes in level of scoring: namely, the absolute amount of time involved in the test, the total number of trials, and the awareness on the part of the subjects that they are participating in a ten-trial, half-hour test. Would the same serial-position effects be observed if: (i) the number of trials were reduced from ten to five but the time allotted for each trial were increased from three to six minutes? (If we did observe the same effects over this five-trial 30-minute test, then sheer passage of time would seem to be more important than number of trials); if (ii) the number of trials were kept at ten but the time allocated for each trial were to be halved to $1\frac{1}{2}$ minutes? (If this same effect occurred over this 15-minute ten-trial test, then time would seem to be a less critical variable than number of trials); and if (iii) the subjects were led to believe that the test would consist of ten trials and extend over half an hour, but at the end of the fifth trial the experimenter were to cut the test short, with the result that the trial which was in fact the final one would presumably have had a different 'mental attitude' associated with it than if the subjects had known that it would be the last? (If (say) the U-shaped curve is due to subjects reacting positively towards the perceived beginning and end trials, then under the conditions described above one might expect that only the first 'half' of the U-curve would appear.)

There are thus a number of interesting manipulations which the researcher could carry out so as to yield more informative data on the critical variable(s) that gives rise to serial-position effects.

(iii) The Sheep-Goat Effect. Far and away the strongest claim for paranormality being made in this thesis (and the claim which if factually correct would constitute the "gold nugget" alluded to earlier) is the suggestion that there was observed fairly consistent evidence of a species of the so-called 'sheep-goat effect', or 'SGE'. Why this statement is qualified with the phrase "a species", is as follows. The term 'sheep-goat effect' was first used by the psychologist Gertrude Schmeidler (e.g. 1945) to describe the relationship between acceptance of the possibility of ESP occurring under the given experimental conditions, and the level of ESP-scoring: those subjects not rejecting this possibility (the 'sheep'), tended to score above chance, those rejecting it (the 'goats'), at or below chance. (Note the narrowness of this definition: Schmeidler did not ask her subjects whether they believed that ESP exists, but only whether it can occur in the given test situation.) Since publication of her pioneering work, a large number of researchers have attempted to confirm this sheep-goat effect. However, the criteria used for classifying subjects as sheep or goats have often differed somewhat from Schmeidler's, and in potentially important ways. Nevertheless, these criteria all have as their common denominator the presence or absence of belief in the occurrence of ESP as either a theoretical possibility or an established fact, with or without supposed

personal experience of such phenomena. Though Schmeidler herself might not approve, this cluster of belief-systems is being increasingly referred to as the 'sheep-goat variable', as if it were a unitary dimension. In fact, in the experiments reported for this dissertation, subjects' belief in, and experience of, cognitive psi, were measured by means of an 11-item Sheep-Goat Scale (hereinafter 'SGS'), whose theoretical range of scores is 0—26, with an empirical mean (for students, at any rate) of about 11.5. The label 'sheep' is thus loosely applied to persons who, compared to the population average, obtain a relatively high SGS score, while 'goats' are those who obtain scores which are relatively low. It is thus a long way from Schmeidler's original dichotomy, especially so when one remembers that the scale also includes items regarding belief in post-mortem survival and in the possibility of communication with spirits of the deceased (which are two variables that have probably not been used on many previous occasions to predict ESP-score).

But be this as it may, the claim that a species of sheep-goat effect was observed is, in detail, the claim that under certain circumstances, the SGS scores of several groups of subjects (whether agent or percipient) have shown significantly positive correlations with the psi-scores mean rank and hits, in the context of a ten-trial free-response test of GESP: the more 'sheepish' the subject, the greater the tendency to psi-hit, and the more 'goatish', the greater the trend towards psi-missing.

Upon discovering such correlations in the data of the Closeness Study (see the introduction to Chapter 3), it was initially thought that the effect was a fairly straightforward one, in that it

would occur in basically unselected pairs of subjects. This hope was very quickly dashed when an attempted replication of the effect, using extreme sheep and equally extreme goats, proved abortive (Chapter 3: the First Sheep-Goat Experiment). The effect again failed to appear in the later experiment with married couples (Chapter 5). To cut short a long story which has been told in detail in the first section of Chapter 4, it became fairly obvious that (strange as it seems at first sight) the SGE was confined to those dyads where the partners were not involved in a mutually close emotional relationship with each other: the effect failed to appear whenever there was such a relationship between the partners (as in the 'close-relaters' of the Closeness Study, and in the married couples), or whenever no attempt was made to ensure that such closeness did not exist (as happened in the First Sheep-Goat Study). Only when the correlation was calculated for that subset of dyads who were 'not-close' was any convincing evidence forthcoming for the SGE.

In the Closeness and the First Sheep-Goat studies, however, the social-emotional situation between the agent and percipient comprising each 'not-close' dyad was confounded with the fact that these subjects were almost always those special sort of people we have called 'non-close-relaters': such persons, who claim never to have had a close, reciprocated emotional relationship with anyone in their lives, tend also to be more shy, more introverted, and more lacking in social skills compared to close-relaters and the population norms (see Thalbourne, 1976; these effects were successfully replicated in a large-scale follow-up study, as yet unpublished).

The experiment reported in the latter half of Chapter 4 was therefore designed with a view to examining whether the critical variable (or constellation of variables) conducing to the SGE was simply the lack of a close emotional relationship with one's partner, or the tendency to possess the distinctive personality characteristics of the inter-personal 'loner'. Pairs of non-close-relaters were compared with pairs of 'mere strangers' (i.e. persons who knew each other only very slightly or even not at all prior to the experiment); these strangers were selected at random from the population of persons who were not non-close-relaters, and were thus (presumably) fairly average personality-wise. Such evidence as was obtained for the SGE (namely, ESP-scoring was correlated with the sheep-goat status of the agent but not of the percipient) showed reasonably unambiguously that 'being a non-close-relater' was not a necessary condition for the production of the effect, seeing as the group of non-close-relaters and the group of 'mere strangers' both gave comparable evidence for it: the SGE seemed not to interact with the personality-syndrome* that typifies the non-close-relater, but only with the presence or absence of a close emotional relationship.

If we combine the data yielded by all 53 'not-close' dyads

---

* The reader may well be curious to know what is the degree of association between the close-relater/non-close-relater dichotomy and these personality dimensions (namely, social introversion, social boldness, social skills, withdrawing tendencies). The author has accordingly, if latterly, computed the values of omega-squared (which is derived from the t-ratio of the group-difference: see p. 100). These values in fact rarely exceed 0.10, indicating that only about 10% of the variance in the personality-measures is accounted for by the close-relationship variable: there is a considerable degree of overlap between close- and non-close-relaters, personality-wise. In hindsight, therefore, one might have thought that the personality-differences found by t-test were not sufficiently great to qualify as a plausible reason for the absence of the SGE in close dyads.

tested in the three relevant studies (namely, Closeness, First and Second Sheep-Goat Experiments), we obtain for the 53 agents a correlation of .52 between the Sheep-Goat Scale and mean rank ($2p = 0.00006$), and for the 53 percipients a correlation of .41 ($2p = 0.002$).

But as Palmer (1975) and Isaacs (personal communication, 1981) have pointed out, the mere fact of a significant correlation does not of itself tell us much about the locus of the ESP-effect, that is to say, whether all subjects are contributing to the effect equally from across the entire spectrum of the sheep-goat dimension: for example, if there is a truncated range of ESP-scores, then it could well be the case that one subject-group (sheep or goats) were displaying psi (either hitting or missing, respectively), while the other group were scoring at chance. Examination of the range of actual ESP-scores gives little indication of such a truncated range: for the mean rank variable, the range for the 53 scores is 4.23 - 6.53 (which are respectively +1.27 and -1.03 points from MCE of 5.50), while for the hits-score the range is 2 - 8 (each of which values is 3 hits from MCE of 5).

But in order to clinch this point, the (eleven-item) Sheep-Goat Scale was dichotomized around the value which on semi-independent grounds seems to be the empirical mean, namely 11.5.* All

---

* This value was obtained for an N of 327, constituting all persons to whom the eleven-item SGS has ever been administered: namely, the 62 Ss from the Closeness Study (mean=11.3), the 60 from the First Sheep-Goat Study (12.1), the 54 from the Second Sheep-Goat Study (11.2), the 28 from the Married Couple Study (12.2), the 16 from Makris (1979; mean=16.8), the 24 from Harding (1980; mean=14.2), and the 83 Ss from Thalbourne et al. (1982; mean=9.3). Of these 327 Ss, 106 are those who constitute the 53 'not-close' agent-percipient pairs (mean=10.1), while the mean of the remaining 221 Ss is 12.1. As further studies make use of the SGS, the estimate of the population parameters should become progressively more accurate.

subjects scoring above this value were classified as 'sheep', all below, as 'goats', and the mean ESP-scores of each group were compared with MCE (separately for agents and percipients) using single-samples t-tests. There turned out to be 20 sheep agents and 33 goat agents. For the mean rank variable, sheep scored significantly above chance (mean = 5.22, deviation = +.28, t = 2.19, 2p = 0.041), while goats scored below chance to a marginally significant degree (mean = 5.71, deviation = -.21, t = 2.03, 2p = 0.051). For the hits variable, the sheep mean was above chance but not significantly so (mean number of hits = 5.45, deviation = +.45), while the goat mean was significantly below (4.36, deviation = -.64, t = 2.56, 2p = 0.016). Thus, there is really no cogent evidence from the agents that sheep and goats contributed markedly different 'quantities' of psi to the significant SGS/ESP correlations.

The story for the percipients (of whom 19 were sheep and 34 goats) is very similar to that for the agents, with neither group producing consistently greater deviations from chance than the other. We therefore seem justified in concluding that the high scale-scorers tended to obtain ESP-scores that were above chance, while the low scale-scorers tended to score below chance on the ESP measures: it is a genuine 'democratic', or 'bilateral' sheep-goat effect. Hence, if these significant results are attributable neither to methodological error nor to statistical artifact, then it would seem reasonable to suspect that they may be due to the operation of ESP. It is high time to turn to the issue of whether the SGE observed here could have come about in ways other than psi.

As we noted in our discussion in the previous section concerning the evidence for serial-position effects, there are several possible methodological flaws that could conceivably account for the ostensible ESP effects: namely, potential sensory-leakage from experimenter to subjects, from agent to percipient, and from experimenter to judge. The author could expend considerable space and ingenuity in arguing, from post hoc examination of the data, that these avenues of sensory communication could not plausibly explain the patterns of significance observed within these data (such as psi-missing on the part of the goats). But even if the reader were satisfied on this score, there might still be the nagging feeling that perhaps the results are due to statistical artifact -- not so much Type I error as the 'multiple analyses' artifact: the more analyses one carries out on the data (this pitfall being particularly difficult to avoid if one has access to a computer, as did the present author), the more likely is one to discover a significant effect which is in fact spurious. It is indeed true that much of the evidence for the SGE in this thesis arises from post hoc analyses (either exploratory, or retrodictive); in only one experiment (the Second Sheep-Goat Study: Chapter 4) was the interactive effect successfully predicted before collection of the data, and even then it held only for the agents and not for the percipients. In order to establish the effect reliably, it is really the case that a further sheep-goat study is required (perhaps deliberately manipulating the close/not-close dichotomy, as well as perhaps the degree of similarity between the sheep-goat status of agent and percipient.)

But let us cut the Gordian knot by proposing the principle that if independent and adequately-controlled studies manage

to confirm the psi-predictive capability of the Sheep-Goat Scale, then this fact will retrospectively cast a more favourable light on the claims being made for the Scale in the present series of experiments: the touchstone of the authenticity of the effects must be that they should replicate in at least some studies outwith this thesis.

Fortunately, there are to date four other studies which have made use of the Sheep-Goat Scale to predict ESP; the manner in which the Scale has fared on these occasions is reviewed at some length in Appendix 23. Each of these four experiments departs in many potentially important ways from the paradigm used by the present author: some differ in mode of ESP tested, some in number of trials given per subject, length of time allowed for each trial, the experimental task, the provision of feedback after each trial, and in the social relationship between agent and percipient. It could have been the case that owing to one or more of these differences, the SGE did not replicate: it might have proven to be the case that the effect occurred only in the context of a ten-trial, GESP free-response task with consecutive, three-minute-long trials without feedback. It turned out, however, that this degree of specificity was not necessary for the occurrence of a significant SGS/ESP correlation: two of the four studies yielded evidence of significant sheep-goat effects, and in one of these (namely, Thalbourne, Beloff and Delanoy, 1982), the task was a forced-choice precognition one! Nor did the SGS show any sign of predicting GESP-score where there was evidence of a close relationship between sender and receiver (Delanoy, 1981). The Scale has thus demonstrated some predictive power even outside

Table 7.1 Correlations between Mean Rank, Hits, and Sheep-Goat Scale and Component Items, for all not-close subjects. (n = 53)

	MEAN RANK		HITS	
	AGENTS	PERCIPIENTS	AGENTS	PERCIPIENTS
Sheep-Goat Scale (11 items)	-.52†	-.41**	+.49***	+.25
Sheep-Goat Scale (10 items)	-.51†	-.40**	+.47***	+.24
Belief in ESP	-.23	-.38**	+.21	+.29*
Belief that one is Psychic	-.31*	-.40**	+.29*	+.24
Personal Experience of ESP	-.43**	-.24	+.36**	+.04
Have had a Hunch	-.24	-.18	+.16	+.11
" " Premonition	-.22	-.20	+.08	+.11
" " Precognitive Dream	-.38**	-.25	+.35*	+.22
" " Vision	+.07	-.14	-.12	+.05
Belief in Life-After-Death	-.31*	-.15	+.30*	+.21
" " Spirit-Contact	-.34*	-.31*	+.38**	+.19
Have had Telepathic Experience	-.53†	-.39**	+.59††	+.29*
Pre-test Estimation	-.30*	-.30*	+.31*	+.21
Post-test Estimation	-.26	-.04	+.43***	+.11

Note: * : $2p \leq 0.05$
** : $2p \leq 0.01$
*** : $2p \leq 0.001$
† : $2p \leq 0.0001$
†† : $2p = 0.00004$ (1 in 240,344)

of the present context. While not all of these four studies yielded confirmation of the SGE, we may note that on all the occasions when significance was found, it was always in the predicted direction, namely with sheep scoring higher than goats; there were no significant reversals of the sheep-goat effect; on Palmer's pattern-analytic approach, we seem justified in concluding that the predictive power of the SGS is real and not illusory. We also seem justified in supposing that the SGE's reported in the present dissertation are more likely to be genuine than artifactual. We may thus now proceed to discuss the meaning and implications of this significant sheep-goat effect.

One of the first pieces of information we might wish to know about the significant SGS/ESP correlations is what are the various contributions that are being made to them by each of the individual sheep-goat items comprising the scale. This information is displayed in Table 7.1. (The coefficients are Pearson in the case of the SGS, and Spearman for the component items. For the sake of interest, the Table also gives the correlations for the ten-item SGS as well (which omits the item Pre-test Estimation of the likelihood of the subject showing ESP), and also for the post-test item "Do you feel that ESP took place during this experiment?" (See Appendix 7)).

We are immediately struck by the fact that all but two of the correlations in this Table indicate 'sheep' scoring higher than 'goats' (although of course this is not so surprising when one considers that all the individual items correlate positively with the Sheep-Goat Scale). The highest correlation is between number of hits scored and the belief on the part of the 53 agents as to whether or not they had had at least one experience of

telepathy: rho = +.59, the anti-chance odds being 240,343 to 1. Ironically, this value is even higher than that for the SGS/ESP correlations! (Indeed this superiority often occurred in each of the three individual studies from which the data in this Table were extracted.) It is apparent that the sort of agent who might be expected to lead to a high ESP-score (given the absence of a close relationship with their partner) is one who claims to have had personal experience of ESP, and of telepathy in particular, to have had at least one precognitive dream, believes there is a good likelihood of them displaying ESP in the test, and who believes both that they are psychic and that post-mortem survival and communication with the deceased are facts. For the percipient the pattern is similar but not nearly so strong; in their case, too, abstract belief in the existence of ESP has a more important place than (say) precognitive dreams and belief in an after-life. In general, then, it might seem that these significant correlations, obtained under reasonably well-controlled laboratory conditions, retrospectively throw a slightly better light on the claims made by those subjects actually to have experienced psi outside the laboratory: if a person does indeed psi-hit in an experimental task, having claimed to have psi-hit previously in the course of their everyday life, then one begins to wonder whether at least some of their claims might not be founded on a factual basis after all!*

None of the above implications would strike a layman as

---

* Persons who claim not to have had psi experience do not score at chance in the laboratory but rather tend to score below chance. What then do they do in their everyday lives? The intriguing implications of significantly negative scoring will be deferred until later, when we consider a possible explanation for the SGE.

being anything more than intuitively obvious and trite. Yet according to Palmer (1978), subjects claiming to have ever had a psychic experience, or believing themselves to be psychic, have very rarely scored significantly higher than other subjects (as in Akolkar & Deshpande, 1966; Moss & Gengerelli, 1968):

> "Such a question was included in the successful composite scales of Bhadra (1966) and Palmer and Miller (1972), but in the latter, at least, its contribution to the scale's success was considerably less than the contribution of the abstract belief question. Beloff and Bate (1970) found that subjects who believed they had psychic ability scored differently from other subjects to a significant degree across seven samples, but the direction of the effect differed from sample to sample. Also, Jones and Feather (1969) found significantly higher between-subject variance among subjects who reported a relatively wide range of psychic experiences. Otherwise, this question has failed to yield significant discriminations .... " (p.159)

Similarly, says Palmer, questions asking how well the subjects themselves think they will score or have scored on the ESP test have very rarely proven significantly accurate (Musso, 1965; and possibly Eisenbud, 1965). One wonders, however, whether if Palmer were to apply his pattern-analytic approach to the list of non-significant studies that he cites he would find nevertheless a trend in favour of the subjects making claims to psi experience. But if Palmer is correct after all, then the effects found in the present series of studies are quite atypical for parapsychology.

One might suppose that if these effects continue to prove replicable (as indeed Appendix 23 suggests), then perhaps it might be appropriate to consider improving the predictive power of the Scale by eliminating items that contribute little or nothing to the overall correlation. For example, Table 7.1 suggests that we might well remove the item on Paranormal Visions and perhaps replace it with the Post-test confidence item. In this way, we could 'build a better mouse-trap' to catch that elusive creature,

psi. A preliminary attempt to do this with the present data, using the seven pre-test items that correlate best with ESP-score in Table 7.1, raised the SGS/hits correlation for agents from +.49 to +.60 (2p = 0.000003, or 1 in one-third of a million). Such a revised scale would of course have to be cross-validated on new sets of data; unfortunately, present indications do not look hopeful: for example, in the data of Makris (1979), having a Paranormal Vision correlated +.87 with ESP-score (n = 8, 2p = 0.005), and Hunches and Premonitions did substantially better than did Experience of Telepathy (whose rho was +.08)! To be safe, then, one should probably stick with the entire scale for the time being.

The next major issue with which we must grapple is the question of the relative importance of the agent and percipient for the SGE. In studies where the agent and percipient are similar in sheep-goat status (e.g. Moss & Gengerelli, 1968), it is not at all apparent whether the score is due to the attitude of sender or of receiver. This problem is particularly germane to the present series of studies, since agent and percipient were each allotted the same ESP-score. We would expect the SGS/ESP correlations to be similar for agents and percipients in proportion to the degree of similarity between the SGS-scores of agents and percipients: as these two latter scores approach an intercorrelation of 1.00, the two correlations with ESP-score approach identity (and of course, non-independence). To what extent, then, do the SGS scores of agents and percipients resemble each other? For the three individual series under consideration, the agent/percipient SGS correlations range from a high of +.86 in the First

Sheep-Goat Experiment (n = 13 pairs, 2p = 0.0002), where like-minded subjects were deliberately paired, down to -.31 in the Second Sheep-Goat Experiment (n = 27, 2p = n.s.), where subjects were paired at random. For the combined sample of 53 pairs in Table 7.1, the correlation between SGS scores is r = +.19 (2p = n.s.). Given the similarity of the two agent/ESP and percipient/ESP correlations (e.g. -.52 and -.41 for mean rank), this low agent-percipient correlation is, on the face of it, somewhat surprising. A way of statistically separating the influence on ESP-score exerted by agents and by percipients is to employ the technique of partial correlation. This technique "provides the researcher with a single measure of association describing the relationship between two variables while adjusting for the effects of one or more additional variables." (Nie et al., 1975, p.302).

When partial correlation was applied to the data of the 53 not-close pairs, it was found that for both agents and percipients the size of the SGS/ESP correlation was reduced, but by such a small amount that there was little change in the level of significance: for the mean rank variable, the correlation for agents was reduced from -.52 to -.49 (50 df, 2p = 0.0002), and for percipients fell from -.41 to -.37 (50 df, 2p = 0.007). Thus, even when controlling for the attitude of the other partner, it is evident that agent and percipient can each exert an influence on the ESP-score that the pair obtains. We may re-iterate the suggestion made at the end of Chapter 4, that it would be an intriguing project to examine what happens to ESP-score when the attitudes of the two parties are diametrically opposite - a situation which has not as yet been the case in any of the three experiments thus far reported.

Let us now turn our attention to the question "What is the modus operandi of the sheep-goat effect?" Most commentators interpret the SGE as suggesting a motivational influence on psi-scoring (e.g. Child & Levi, 1979). Probably the most plausible hypothesis encountered by the present author is John Palmer's (1972) conjecture that the SGE results from the subjects' "need for vindication", or "the need to defend the validity of one's previously formed opinions on important ... issues" (p.10). Disagreement, says Palmer, "particularly if it comes from a "competent" person, is threatening because it implies the possibility that S is incompetent in the sense that he failed to reach the "correct" conclusion on the issue. He therefore seeks to defend his opinion as a means of defending or vindicating his competence. This conceivably could be a factor in the motivation of sheep and goats to score above and below (or at) chance, respectively, on ESP tests, because their scores provide a sort of evidence about the validity of their previously formed opinions." (p.10) (Failure to guess the targets correctly seems to many goats to prove their thesis that ESP does not exist.)

The present author's own interpretation of the SGE is very similar to Palmer's, except that it is based on Leon Festinger's (1957) theory of Cognitive Dissonance (dissonance being the cognitive discomfort produced by awareness of two inconsistent cognitions). The hypothesis being proposed is that psi in general frequently operates in such a way as to reduce cognitive dissonance -- to make the world appear to the subject to be consistent with his own beliefs as to how it should be and/or consistent with

his desires as to how he wants the world to be.* The same idea could perhaps be expressed in Kuhnian terms (Kuhn, 1970) by postulating that psi operates to promote events that give the appearance of being consistent with a person's own 'paradigm', or view of the world, and to eliminate events which are anomalous on that same paradigm.

What sort of evidence leads the author to believe that such a process accounts for the SGE? In the case of sheep, it is quite obvious that the attainment of above-chance scores is consistent with their beliefs about themselves and the way the world works. Achieving above-chance scores would arouse dissonance, however, in goats, who believe and want such a thing not to be possible. But what leads the goats to psi-miss rather than to score at chance? The author here is drawing upon his own personal observations of the way in which people react to an instance of significant negative scoring. The fact of the matter is that statistically naive persons do not recognize that significant negative scoring is actually evidence for psi (albeit psi-missing). Psi-missing is initially a very counter-intuitive concept, the understanding of which requires a measure of statistical sophistication. Even some psychologists, if they are skeptics, are inclined to argue that the concept of psi-missing is a step on

---

* A not dissimilar Festingerian approach was taken by Stanford (1964) in his attempts to explain the differential position effects he observed for above-chance scoring sheep and goats. There is also some affinity between the present author's hypothesis and Stanford's concept of 'conformance behaviour' (Stanford, 1978); this concept refers to the reorganizing of a relatively random, or unordered, system so that its new state more readily subserves the 'disposition' or goals of another relatively structured, or ordered, system. Space does not permit us to explore here the points of contact and of departure between the two hypotheses.

the road to accepting any significant deviation from chance as being indicative of a real phenomenon; they say that parapsychologists are "having their cake and eating it too", and are "trying to win both ways"; such persons overlook the strength of the evidence for psi-missing (see for example, Schmidt, 1969) and also forget that admission of the possibility of psi-missing entails the necessity of using a two-tailed test of significance (in the case of a single sample of scores), which automatically demands an even more extreme deviation from chance than does a one-tailed hypothesis (i.e. if we are only prepared to allow significantly above-chance scores).

Subjects, too, are equally skeptical about the implications of significantly negative scores. For instance, in the author's Closeness Study, one engaged couple was tested and obtained a significantly negative score. Later, when it came time to solicit subjects for the Married Couple Study, the experimenter was most eager to re-test this pair, who had by now married; but nothing could prevail upon them to be retested: they refused, politely, on the grounds that they had failed miserably the first time! (The husband was a sheep, the wife, a goat.) Subjects naively feel that the true zero-point (i.e. "no ESP") is, in forced-choice tests, "no correct guesses at all", and in free-response tests, "utterly no perceptible similarity at all between target and response": the more correct guesses (or target-response resemblances), the more ESP, they think. (So much for the recent claim that belief in the occult is the result of widespread deficiencies in reasoning and limited knowledge of the laws of probability (Singer & Benassi, 1981; Benassi et al. 1981); it may well be that disbelief in psi is occasionally due to these self-same

deficiencies!) It then follows quite naturally that naive goats should tend to be motivated towards obtaining the lowest score possible, so that "chance" may vindicate their belief that psi does not occur.

If this model is correct, then there is a set of testable predictions that may be made. Suppose we had a group of goats who upon being tested for psi gave evidence of significant negative deviations. A psychological prediction is that these subjects, when informed about these results, will report that they conform to what they would have expected, and that they will show incredulity if an attempt is made to argue that they have actually demonstrated psi. Next, suppose that half of the subjects are taken aside and instructed by a charismatic statistician to believe that their scores were a very rare event according to the laws of probability, and that therefore either a Type I error or psi is indicated. The present author predicts that, if these educated goats begin to worry that their scoring may well have been due to psi, then the consequent dissonance will motivate them to want to be rid of the anomaly and its unpalatable conclusion: It is further predicted that if all subjects are again tested, the non-sophisticated goats will tend to persist in their significantly negative scoring, but the sophisticated goats will give scores that are at a level close to chance (with perhaps even significantly low variance). Such a regression to chance could then be plausibly attributed to the need for the subjects to produce an outcome which is consistent with the way they need the world to be if they are to avoid cognitive dissonance; psi would have been used to suppress the appearance of anomalies

which contradict a person's own paradigm of reality.

If such an experiment ever confirmed these predictions, it would have momentous implications for parapsychology. The evidence provided by this thesis already suggests that a distant person (namely, the agent) can influence the outcome of an event (namely, the drawing of a response by the percipient) in accordance with his sheep-goat attitude. (This will be further elaborated below.) Could it be the case that experimenters can, on occasion, unwittingly act the part of agent and thereby influence the outcome of their own research? Could trans-experimental declines be causally related to changes in the sheep-goat motivation of the experimenter, as the initial desire for "significant results", when fulfilled, begins to give way to ambivalence (could it be Type I error? Is there a flaw in the methodology somewhere? Can it be replicated?), or even to uneasiness or threat (what sort of dynamite are we playing with? Am I immune from the evil thoughts of others?) as the prospect of further significant results turns into a spectre? Furthermore, could not this teleological effect be exerted by other researchers in the same profession, or even by non-parapsychologists and opponents of psi, when brought face to face with a claim for significant results? Could not the problem of repeatability (which is not even unique to psychical research) have come about, and be persisting, as a result of some highly skeptical audience (experimenter, colleagues, professional peers) 'sabotaging' present and future research by needing and hoping that non-significant results will ensue, in order to remove the threat to the Paradigm? Could C.E.M. Hansel really be an outstanding psychic?! Could Thouless' 'witness effect'

be a special case of the sheep-goat effect? Is the sheep-goat effect better termed an 'audience effect', that audience being the subject, the experimenter, or outside observers? Whence, then, is the psi coming in our experiments? Who, in the audience, determines what the outcome of the event will be? Such is the Pandora's box that would be opened if we could establish the existence of pure teleological causation, in which a person exerts prospectively an effect on a present or future event so that it will conform to his or her needs and desires!

But let us come down from these vertiginous heights of speculation and return for a moment to the mundane world of data. Most of these speculations rest on the assumption that the 'agent' may paranormally influence an event in accordance with their sheep-goat attitude regarding the desirability of that event. The significant correlations (partial and zero-order) between ESP-score and the Sheep-Goat Scale-scores of the agents observed in this series of experiments, were the data upon which this assumption was made. Let us focus on the meaning of these findings.

These results immediately suggest that not only the percipient but also the agent can be causally efficacious under GESP conditions. While the agent may not perhaps be a "vital cog" in the process of the percipient's paranormally acquiring information, nevertheless they are not simply so much "excess baggage" (this witty antithesis being used by Palmer, 1978, p.96). These findings are therefore very relevant to the ongoing controversy over whether there really is such a thing as telepathy in addition to clairvoyance. As we shall see below, the findings still do not bring us much closer to establishing the fact of telepathy as mind-to-mind 'thought-transference', but they do nonetheless put the

role of agent into a new light. It is very telling that the Rhinean term for 'percipient' even in a GESP test is often simply 'subject', as if the agent were simply a stage-prop and the mode of ESP 'merely' clairvoyance on the part of the percipient (to which viewpoint Rhine himself was very strongly inclined: see Rhine, 1974). The agent has been the second-class citizen of parapsychology. It is not surprising, then, that in Palmer's famous review of 24 sheep-goat studies (1971), not one of the five which used a GESP procedure attempted to predict the ESP-score of the pair by means of the agent's attitude. The results of this thesis, by affording evidence for what we may call an "Active-Agent Effect", suggest that the sender has been unjustly ignored, and that redressing of the balance is in order: it is one of the few studies to have granted the agent "equal rights".* There are, perhaps, only two other experiments where this has been the case. According

---

* Proponents of the view that the basic process in ESP is clairvoyance (the effective target being a physical object such as a drawing, rather than the agent's mental awareness of the drawing (or even the neural events correlated with this awareness)), argue that the agent has an influence only insofar as he or she affects the attitude of the percipient (subjects commonly believe that having an agent in fact helps them to acquire the target-information (see, for example, Musso & Granero, 1973)). It should be pointed out that since the agents and percipients who showed the sheep-goat effect tended also to be strangers to one another, interacting only briefly before the actual ESP test, it is probably not the case that the correlations with ESP-score came about as a result of the percipient's knowledge of the attitude of the agent and thereby conforming to what is expected of him, by psi-hitting or psi-missing. If this argument can be sustained, then the data of this series of experiments seriously embarrass the 'clairvoyance' school of thought.

to a recent report by Eisenberg & Donderi (1979), an attempt was made to correlate GESP-scores with the personality and attitude of the agents, but without, it must be admitted, conspicuous success. However, a rather more successful experiment was reported twenty years ago by Schmeidler (1961), and her results well repay close study, since they offer a precedent, and indeed a useful explanation, for the Active-Agent Effect.

All Schmeidler's percipients believed that the card-tests in which they were participating were being run under GESP conditions; unbeknownst to them, however, three variations were introduced in the sending conditions: in the first, the agent simultaneously viewed the target-card (as per the classical GESP paradigm); in the second, the 'Success' condition, the agent was ignorant of the target but hoped the percipient would succeed in the task of acquiring information about it; and in the third, the 'Failure' condition, the agent was again ignorant of the target but hoped the percipient would fail in the ESP-task; these latter two conditions would conventionally be described as 'clairvoyance' tests.

The ESP-scores obtained under the three conditions were compared using analysis of variance, correlation coefficients, and tests for the difference between two correlations. Firstly, there was a highly significant difference between the mean scores for the GESP, Success and Failure conditions. Secondly, there was a significant correlation of -.20 between GESP and Failure scores, and a marginally significant difference between the two correlations for GESP/Success and GESP/Failure; these findings led Schmeidler to conclude that "telepathic messages of encouragement or discouragement, without specific content, have different

effects on the percipient" (p.23). There was also a significant difference between the two correlations for Failure/GESP and Failure/Success, which suggested that something different happens when the agent hopes for success knowing what the target-message is, and when they hope for success without knowing the specific piece of information to transmit. These results together led Schmeidler to infer that the agent does have an effect, and that this effect can be either the communication (or perhaps the 'amplification') of a specific signal, or, even without the agent knowing the target, the causing of a change in the percipient's scoring.

Schmeidler may thus have been one of the first researchers to find evidence of an Active-Agent Effect, albeit not a classical sheep-goat effect; even so, perhaps the Success and Failure conditions are not so very different from what sheep and goats feel in their heart of hearts that they would like to be the case.

Now while it does not seem so very implausible that, in the present series of experiments, the sheep-goat status of the percipient should be related to the degree to which the drawing he or she makes resembles the distant target, it is rather more startling (or at least would have been, were it not for the results of Schmeidler's (1961) study), that the attitude of the agent should be as determinative if not more so of the degree of target-response similarity. For in the case of the percipients, we can imagine the psi-process as being analogous to the scanning of the environment by radar: in the case of a sheep percipient, the information is correctly acquired, and manages to manifest itself on paper as a response resembling the target; in the case of a goat percipient, the correct information is still apprehended

at some level, but for whatever reason (and we have conjectured that this 'reason' is the desire to vindicate chance by producing a 'low' level of resemblance), the information that finds expression in the response is quite the opposite of what is in the target. (The selection of a topic which will constitute an antithesis of the target-information, is almost certainly done below the level of conscious awareness, though it would be interesting to have some introspective data from percipients on this point.)

But what is happening in the case of the agents? Presumably, one could imagine a sheep agent successfully 'transmitting' the correct target-information to the percipient, the latter somewhat passively reproducing it on paper. But what of the goat agent? How does he or she bring it to pass that the percipient's response shows significant avoidance of the target-drawing's content? On an 'information-transference' model, one would have to suppose that the goat agent were perverse enough to try to send a message which was diametrically opposite, in concept and form, that portrayed in the target! This seems unlikely in the extreme; indeed, the present author cannot remember a single instance of it happening that when he and the agent and percipient were going through the responses and their targets together at the completion of the ESP-test, the (goat) agent reported that though the responses do not match the targets they do consistently match various thoughts that he or she was having at the time (these thoughts also being the very antithesis of the target-theme). It is much more as if the agent said to themselves "I want the response-drawings to show a low degree of resemblance to these targets" and that desire somehow became actualized in reality.

By what psi-process, then, does this effect occur, if not

by paranormal transmission of information? Gertrude Schmeidler (1961) has some very apropos remarks, and with which the present author fully concurs. She points out firstly that:

> "We have many phrases [in everyday language] which imply the belief that wishing a person well or ill, rooting for him or against him, giving him a blessing or a curse will help determine the course of events in some way independent of his knowledge of these wishes, and similarly independent of the wisher's knowledge of the precise events to be affected." (p.36).

She then suggests an analogy which, without actually explaining the Active-Agent Effect, nevertheless brings it into congruence with other similar data:

> "There is, it seems to me, a parallel between the apparent effectiveness of Success and Failure wishes and PK with dice. It has often been pointed out that one of the fantastic things about extrachance effects with die faces is that the dice move too rapidly for the subject to know what is happening. If he wants to give the dice a psychic push so that, let us say, the 4-face is up, he has no time before the dice have stopped to decide which way to push them. Thus, though the data indicate his wishes can be effective, he does not know what has made them so." (p.37).

If psychokinesis can operate thus in a goal-oriented fashion, and, apparently, regardless of the complexity of the target-system (see Schmidt & Pantas, 1972; Schmidt, 1974), then perhaps it can account for the Active-Agent Effect as well, and the latter is not after all anything remotely resembling information-transmission! This is not an entirely new idea: F.W.H. Myers (1903, pp.195-196) pointed to the possibility of a direct PK action of one person's thought upon the brain and nervous system of another, as is suggested in the old 'Willing Game'; he even named this phenomenon "telergy". Thouless & Wiesner (1947) also recognized the possibility of a psychokinetic form of telepathy (as opposed to the true cognitive telepathy, in which the percipient acquires information directly about the agent's mind or brain, this being

a species of clairvoyance): they called the former type " κ-telepathy" (from "psi-kappa" -- their term for PK), as opposed to " γ-telepathy" (from "psi-gamma": ESP). Though the analogy has frightening implications, we might suggest that, by willingly submitting to the role of percipient, a subject exposes him- or herself to psychic manipulation by the agent, as if the latter were a puppeteer and the former a marionette.

Schmeidler (1961) also goes on to remark that

> "PK theorists are likely to make their next step, in discussion of this mystery [concerning the dice-effects], an affirmation of similar ignorance for ordinary behavior. We do not know which muscles must be tensed and which relaxed in such skilled movements as writing or tying shoelaces, and yet the necessary movements usually occur in accord with our wishes. And many college lecturers will testify that we open our mouths with a vague notion of what we want to say, and are interested and sometimes astonished at the words that come out, though here again our words have a general tendency to conform with our intentions. The difficulty of target ignorance thus takes its place in a familiar company of comparable difficulties." (p.37)

Schmeidler might also have cited, in addition to such mysterious processes as speech-production, the phenomena of psychosomatic disease, and perhaps hypnosis.

In view of these similarities, it may be useful to apply the principle of parsimony and, like Thouless & Wiesner, suggest that the only difference between these somatic processes and psychokinetic ones is their location -- whether they are endosomatic or exosomatic. PK, then, would simply be the exosomatic form of processes that are normally endosomatic; or alternatively, one could say that endosomatic processes are brought about by PK.

Thouless & Wiesner also postulate an identity for the processes of sensory and of extrasensory perception, the former being a result of clairvoyance of one's own brain-states. The present author would suggest we go even further, and postulate that this

same process extends to all cases of information-acquisition, whether paranormal or normal, the latter category encompassing the processes of retrieval of information from memory, creativity, and other forms of mental problem-solving. The Self may well be in a relation to these normal modes of information-acquisition not so very different from paranormal modes of obtaining knowledge.

Though Thouless & Wiesner say at one stage that psi-action and psi-cognition may each be simply different aspects of the same unitary process -- linked by properties other than the fact that they are both paranormal -- they nevertheless go no further in attempting to delineate an underlying basis for the identity, and indeed persist in discussing 'psi-gamma' and 'psi-kappa' as if they were different 'forces'. The author would like to go one step further and postulate such a basis for identity of ESP and PK (and hence, also, of our control over our bodies and our minds). The suggestion is that the denominator common to all these processes is the achievement of some goal (i.e. a wish or a need): in the case of ESP, or (say) memory, the goal achieved is the acquisition of need- or desire-relevant information, while in the case of PK, or the volitional control of the body, the goal is manipulation of the physical environment. The fact of goal-orientation provides the identity; it is only a contingent and superficial difference that in the one case fulfillment is achieved on the mental level, and in the other, the physical realm. The Self, then, is thus conceived as being in the same relation to its mind as it is to its body; it expresses the desire, and under appropriate psychophysiological conditions (e.g. the person is alert rather than tired, sober rather than drunk, mobile

rather than paralyzed), the mind/body complex responds, automatically, and (we might say) magically, without our knowing the precise events to be affected in order to bring about the effect. Mind/Body Dualism is thus replaced by Self/non-self dualism. In this way, we may achieve unification across both normal and paranormal processes, and across action and cognition. In a sense, it is all action, but the action of a Master bidding a more or less responsive slave to carry out his orders.

In order to describe this unitary process, which transcends psychology and parapsychology and yet which weds them together, the author has coined the term "psychopraxia" (from the Greek psyche, meaning 'soul', and 'praxia', derived from prattein, 'to accomplish, bring about'). 'Psychopraxia' may be defined as "a fundamental teleological principle underlying all Self-to-mind and Self-to-matter interactions, whereby, under certain (as yet unspecified) psychophysiological conditions, the adoption of a desire automatically results in its fulfillment in reality". Psi is seen as a special instance of psychopraxia, being those manifestations which are exosomatic rather than endosomatic, that is, which are not mediated by the body's cognitive-motor apparatus. Thus 'psi' may be said to be occurring quite normally within our bodies, with such efficiency that we take it for granted without appreciating the basic mystery of how it happens; one might call this most elementary level of psychopraxia 'alpha psi', which babies and children learn to master at an early age.

Though space does not permit of a fuller discussion, the concept of psychopraxia may be particularly useful for cases of psi which cannot be neatly pigeonholed as being definitely PK or ESP, such as the Psychic Shuffle, and Psi-dexterity tasks

(Knowles, 1968): we can then say that psi operates simply so as to tend towards the fulfillment of some goal, whether that can be accomplished in detail as a result of ESP or PK or a combination of both. It is perhaps not out of place to point out that even in cases of normal motor skills, for example playing the piano or writing a dissertation, the production of the effect depends on a complex interaction between sensory and motor activities, in a cybernetic feedback loop; even in trying to recall some piece of information, we often engage in some physical action (such as closing our eyes) to facilitate the bringing about of the effect.

If we now wed the concept of psychopraxia to the dissonance-reduction process mentioned earlier, we arrive at the conception of a process which is a cybernetic one, in which the Self strives to maintain the goal of cognitive homoeostasis. The fact that the World gives some semblance of stability on the macroscopic level, and is not a Chaos of elements each attempting to fulfill competing and incompatible desires, is testimony to the infrequent occurrence of psi (or rather, of exosomatic psychopraxia): the Sheep-Goat Effect, for example, is a very weak effect indeed. Yet if we start with the admittedly wild assumption that a human being can, under certain as yet unknown (but probably psychophysiological) conditions attain what we may call 'omega-psi' -- the ultimate end of psychic development -- in which he or she has access to omniscience and omnipotence, then, as Beloff (1973b) points out, the question "is no longer the familiar one, i.e. "How does it work?", but rather "Why, if this [state of omega psi] is the case, is [psi] so rare, so unstable and unreliable, so inaccurate and, for the most part, so trivial in content?" (p.299) Why cannot we all be thaumaturges, like Jesus, or Socrates, or perhaps an Apollonius of Tyana or even a Sai Baba? What are

the conditions that impede more voluntary and more powerful use of psi? This, the author believes, is the question which it is the principal business of parapsychologists to try to answer.

Beloff himself hints at some of the conditions that seem to accompany outstanding performances of ESP: "Such achievements would seem to demand an almost superhuman power of concentration and a unique state of mind compounded of extreme passivity towards incoming impressions with an ardent striving to know the answer." (1973,b; p.303). Curious, that this description is not unlike phenomenological instructions for retrieving a particularly elusive piece of memory-information!

Perhaps, in the present series of experiments, some condition was present in the close dyads that militated <u>against</u> the manifestation of a sheep-goat effect, and which was a condition <u>not</u> present in the <u>non</u>-close agent-percipient pairs. The author suspects that this factor was something like 'over-anxiousness to psi-hit' -- psi-hitting being perceived as necessary for establishing that the relationship was 'psychically' as well as emotionally close, or else necessary because it was <u>expected</u> of them as per the Closeness Hypothesis. The more blasé attitude of two strangers thrown together and asked to attempt to communicate telepathically, might have been of just the right character to present an interesting challenge to the subjects' need for vindication without at the same time them feeling oppressed by the <u>need</u> to psi-hit. There is indeed evidence that where there was anxiety or lack of confidence about succeeding (i.e. psi-hitting), the SGE did not occur: the 36 close-relaters of the author's first study turned out to be significantly more 'agitated' during the ESP test, by their own report, than the 26 non-close-relaters (2p

= 0.044) (Thalbourne, 1976, p.55); the non-close pairs manifested the SGE, the close pairs did not. Again, in the Second Sheep-Goat Experiment (Chapter 4), significant correlations were found between ESP-score and the SGS-scores of the 27 agents but not of the 27 percipients, and the latter reported being significantly more agitated than the former. Thus, the lack of a close relationship may not after all be so important as a variable frequently confounded with it, namely, lack of a dire need to prove oneself as a psi-hitter in the context: failure at the ESP task would not be perceived as particulary upsetting or even surprising (given that few people credit strangers with the capacity to communicate by psi), and this absence of threat may provide just the right sort of open attitude to the results that (paradoxically) leads to actual psi-success, which is to say, the sheep-goat effect.

Thus, a necessary condition that must obtain for the successful manifestation of psychopraxia, might, it is suggested, be the absence of excessive anxiety over succeeding. It is a well-known truism that if one tries too hard one may well end up failing, anxiety beyond a certain optimal level being disruptive of skilled performance. Only by removing such anxieties from members of close pairs, could one have any chance of observing a sheep-goat effect. (This could perhaps be achieved by leading two close-relaters who were involved with each other to believe that they were taking part in a telepathy test with a stranger, while all the time they were actually being tested with each other: the closeness would still be there, but the expectancies of success might be reduced to a less anxiety-arousing level, thus allowing the sheep-goat effect to occur). Indeed, only by removing excessive

anxiety can human achievement in all contexts be optimized. Perhaps this is why anxiety-reducing techniques such as Transcendental Meditation, are anecdotally reported by their practitioners to result in greater fulfillment on all levels of achievement, and also in more frequent experiences of intuitive knowledge.

A P P E N D I C E S

APPENDIX 1. TARGETS

SET 1

(5273) bucket
(9229 gauntlet
(5588) storm-lantern
(8218) bee
(6562) padlock
(0224) vase
(1403) farm-house
(4044) piggy-bank
(2487) bust (statue)
(1632) car

SET 2

(7646) spoon
(6479) owl
(8727) ash-tray
(5524) teepee
(3650) globe (terrestrial)
(3093) whale
(4236) revolver
(9904) rocking-horse
(9357) dollar
(4332) log (tree)

SET 3

(1943) square (figure)
(0457) boot
(7091) sofa
(4731) cow
(4223) junk (ship)
(1290) machete
(9553) belfry
(3401) fish
(9449) toucan
(7264) wine-glass

SET 4

(0045) waste-paper basket
(5743) salt (shaker)
(4166) stamp (postage)
(8237) column (Grecian)
(1535) volcano
(9430) nonagon
(7303) pipe (smoker's)
(9875) sombrero
(5448) lorry
(1551) dumbells

SET 5

(1200) razor
(1879) coin
(5238) drum
(4822) clamp
(7167) eye-ball
(2281) watch
(2058) coffin
(5239) swan
(7190) hod
(6660) island

SET 6

(3311) mail-box
(1950) surplice
(0974) train
(2344) shell
(1219) dowsing rod
(1193) cube
(6060) sock
(6138) leaf
(2502) euphonium
(0047) escutcheon

TARGETS [continued]

SET 7

(1864) lute
(8355) pie
(9807) ladder
(0316) bell
(1114) crook
(3323) ghost
(9845) big-top
(8576) shooting-star
(9403) ellipse
(5468) incandescent lamp

SET 8

(6435) telephone
(6189) label
(6370) Pegasus
(5925) tie
(4241) spark-plug
(8519) lyre
(8185) spiral
(5272) light-house
(3291) chalice
(1893) carrot

SET 9

(5865) nail (iron)
(9815) scissors
(8756) anchor
(9340) mountains
(3824) nail (finger)
(0641) gallows
(7111) front door
(2501) wind-vane
(0661) television set
(8129) tyrannosaurus rex.

SET 10

(6875) aqueduct
(1780) lucky dip
(6293) lunar module
(5491) vacuum cleaner
(0596) elephant
(4485) butterfly
(2016) inner tube
(9770) snake
(6226) mortar (bowl)
(7728) book

Set 11

(8969) asterisk
(9119) fan
(3329) print (foot)
(9859) saw
(7528) hamburger
(5900) aeroplane
(7403) funnel
(7025) rectangle
(1703) buttocks
(4743) superphosphate

SET 12

(5299) bat (animal)
(2367) tabelberg
(6934) parabola
(7682) dynamite
(5369) Siamese twins
(5453) hendecagon
(8955) volcano
(6823) stump (tree)
(2633) dam
(9742) toilet-paper

TARGETS [continued]

SET13

(3560) trophy
(9855) radiator
(6770) ear
(6254) banner
(4498) Saturn
(8310) insecticide
(3350) clog
(6931) soft-drink
(0945) clam
(6554) teddy bear

SET 14

(1707) faucet
(3843) tray
(8964) cactus
(3952) dressing-table
(4874) beetle
(8366) coronet
(6136) smoothing iron
(3964) stethoscope
(9396) fire extinguisher
(8134) diaper

SET 15

(9835) heart
(5221) brolly
(4910) medicine-ball
(8701) ale
(3054) stumps
(7506) decoration (medal)
(3353) plum-pudding
(5373) coal
(4859) mannequin
(8311) row-boat

SET 16

(5092) christcross
(3899) gem
(6112) sunblind
(3220) lollipop
(4928) lipstick
(5426) fly (insect)
(9152) stadium
(0537) hamper
(7508) clinical thermometer
(1166) loo

SET 17

(4559) analgesic
(2810) wagon
(5579) ladle
(3301) barber's pole
(8609) pennant
(4453) adhesive
(6414) spear
(8642) key signature
(4571) string
(9759) birdbath

SET 18

(0222) pram
(6429) duomo
(0927) jetty
(2858) Australia
(9893) cauldron
(5438) toffee
(5168) orange
(0146) barrel
(1895) mop
(9679) doghouse

APPENDIX 2

The Closeness Questionnaire

Consider the following questions carefully, and circle the answer that you feel is most applicable. It is important that you answer as truthfully as you can. All information will be kept strictly confidential, and is used merely to compile statistics.

(1) Is there a person with whom you have a very special relationship, such that you feel extremely close to them on a personal and emotional level ?

YES/NO

If "YES" to (1):

(2) Do you think that that person feels much the same way towards you ?

YES/NO

If "NO" to (1) or (2):

(3) Has there ever been a person with whom you have had a very special relationship such that you felt extremely close to them on a personal and emotional level ?

YES/NO

If "YES" to (3):

(4) Do you think that at the time, that person felt much the same way towards you ?

YES/NO

NOTE: A subject is classified as a 'non-close-relater' if they say "no" to both question (1) and question (3). A 'close-relater' is a subject who says "yes" to both questions (1) and (2).

APPENDIX 3

Subject No:_____

PRE-EXPERIMENTAL QUESTIONNAIRE

(1)	I believe in the existence of ESP.	(a) true, (b) uncertain, (c) false,
(2)	I believe I have had personal experience of ESP.	(a) true, (b) uncertain, (c) false.
(3)	I believe I am psychic,	(a) true, (b) uncertain, (c) false.
(4)	I have had at least one hunch that turned out to be correct and which (I believe) was not just a coincidence.	(a) true, (b) uncertain, (c) false.
(5)	I have had at least one premonition about the future that came true and which (I believe) was not just a coincidence.	(a) true, (b) uncertain, (c) false.
(6)	I have dreamt at least one dream that came true and which (I believe) was not just a coincidence.	(a) true, (b) uncertain, (c) false.
(7)	I have had at least one vision that was not an hallucination and from which I received information that I could not have otherwise gained at that time and place.	(a) true, (b) uncertain, (c) false.
(8)	I believe in life after death.	(a) true, (b) uncertain, (c) false.
(9)	I believe that some people can contact spirits of the dead.	(a) true, (b) uncertain, (c) false.
(10)	I believe I have had at least one experience of telepathy between myself and another person.	(a) true (b) uncertain, (c) false.
(11)	How likely do you think it is that you and your partner will display ESP in this experiment?	(a) no likelihood at all, (b) very unlikely, (c) unlikely, (d) 50:50 (e) likely, (f) very likely, (g) certain.

Note: In order to obtain a subject's Sheep-Goat Scale score, award 2 points for a "true" response, 1 point for "uncertain", and none for "false"; for item 11), award no points for a), 1 for b), and so on up to 6 for g). The scale-score is then obtained by summing the point scores.

APPENDIX 4

TELEPATHY-TEST SUMMARY SHEET (SG2STUDY)

Date: _____ Time: _____

Pair: _____ _____ NCR/S

Sender: _____

Receiver: _____

	TARGET	RESPONSE
TRIAL	SET:	SET:
1st		
2nd		
3rd		
4th		
5th		
6th		
7th		
8th		
9th		
10th		

Order checked?

APPENDIX 5

INSTRUCTIONS TO SENDER:

On your left is a box containing a total of ten pictures, face down. Your task will be to concentrate on each picture in turn, in an attempt to send an image of it telepathically to your partner in the next room.

You will also notice a red light. The moment the red light goes on, you will hear a tone, and together they are the signal for you to take the topmost card out of the box and turn it over, so that you can see the drawing depicted. The red light will remain on for 2 minutes, in which time you will try to send a telepathic image to your partner (see below).

The moment the red light goes off, the tone will sound, and together they are the signal for you to stop trying to transmit the picture, and to place the card, face down, in the box on your right. The light will remain off for about a minute. It is suggested that during this rest-interval you should try to make your mind a complete blank: forget about the drawing which you have just been looking at, and dismiss ALL thoughts from your mind until the red light goes on again, ready for the next picture.

Throughout the experiment, it is important that you remain quiet and at ease. Make yourself comfortable, and allow your tensions to fall away. Set aside any worries or distracting thoughts. Saturate yourself with confidence, with enthusiasm, and with a desire to succeed.

INSTRUCTIONS FOR TRANSMITTING THOUGHT:

Based on the findings of previous psychical researchers, several methods have been suggested for increasing the likelihood that ESP will occur. Since we want to maximize the likelihood of ESP taking place in this experiment, would you therefore please, during each 2-minute sending-time, spend some time doing the following things:

(1) Concentrate on the drawing and mentally trace the lines of the drawing, over and over again,

(2) Become passive. Allow your thoughts and feelings to centre around your partner. Then, as vividly as you can, visualize the face of your partner: imagine in your mind that he or she is sitting in front of you. Next, as though you were in your partner's actual physical presence, mentally describe the picture you are trying to communicate.

Go on repeating methods (1) and (2) until the light goes off.

APPENDIX 6

INSTRUCTIONS TO RECEIVER:

In the adjacent room, your partner has a series of ten pictures. He or she will concentrate on each picture in turn, and your task will be to try to draw the thing that is depicted on each card. The pictures are not complicated scenes: they are drawings of objects or living creatures; some are very simple, while others are more elaborate. Anyone of average drawing-ability should be able to produce a recognizable copy of the drawing within two minutes.

On your left is a box containing a total of 10 sheets of drawing paper, each code-numbered. You will also notice a red light. When this lights up, you will hear a tone sound, and together they are the signal that your partner is trying to send you a mental message about the picture at which he is looking. So when the light goes on and the tone sounds, take the topmost piece of paper from the left-hand box, turn it over so that the numbered side is facing downwards, and have your drawing-pen ready. You will have a maximum time-limit of 3 minutes per drawing in which to try to reproduce the target-picture. Note well, however, that the red light will remain lit up for 2 minutes only, per picture; it will then go out for 1 minute before coming on (with the tone) to signal that the next picture is being transmitted.

It is suggested that when the light goes on and the tone sounds, you remain relaxed but attentive. Let your thoughts be about your partner. Pause about a minute before attempting to draw what you think the target-picture is. If you get some kind of feeling or hunch while the red light is still on, that's fine. You may find, however, that you still have no impression at all by the time the light goes out. In that case, just draw whatever you like, that is, whatever comes into your mind: you will still have a minute in which to draw your picture before the red light comes on again.

IT IS ABSOLUTELY VITAL THAT YOU DRAW SOMETHING ON EACH OF THE 10 SHEETS OF PAPER. You may, if you like, label each drawing, saying what object or living thing you have drawn. (Note also that you may, if you really want to, draw an additional object on the one piece of paper: you may get a strong hunch, after you have already drawn something, that the target-picture is really of something else).

When the red light comes on again and the tone sounds, place your drawing face down in the right-hand box, even if you haven't quite finished drawing it. Take the next blank sheet from the left-hand box, turn it over, and repeat the procedure described above.

APPENDIX 6 continued

NOTE WELL: You may have heard someone (who has already been through the experiment) talking about what the pictures were of. If so, please bear in mind that the 10 pictures that are going to be used in this particular session of the experiment have been drawn at random from a very much larger pool of pictures. This means that whatever you have heard about the pictures, it will be of no use whatsoever in helping you decide what the pictures are in this session.

Throughout the experiment, it is important that you remain quiet and at ease. Make yourself comfortable, and allow your tensions to fall away. Set aside any worries or distracting thoughts. Saturate yourself with confidence, with enthusiasm, and with a desire to succeed.

APPENDIX 7

Subject No: _____

POST-EXPERIMENTAL QUESTIONNAIRE

(1) (i) Do you feel that ESP took place during this experiment?
(a) yes
(b) uncertain
(c) no

(ii) Do you think that there was anything about the experiment that made ESP less likely to occur?
(a) yes
(b) uncertain
(c) no

(iii) If you answered 'yes' to (ii) what do you think it was that made ESP less likely to occur?

.....

.....

(2) These are several rating-scales to help you describe how you felt during the telepathy experiment. Place a cross (X) at the point which best describes how you felt: you may put the cross anywhere at all on the scale. Do this for all five scales.

(i) tense :____:____:____:____:____:____: relaxed

(ii) bored :____:____:____:____:____:____: interested

(iii) calm :____:____:____:____:____:____: agitated

(iv) confident :____:____:____:____:____:____: uncertain

(v) unenthusiastic :____:____:____:____:____:____: enthusiastic

APPENDIX 8

Targets and responses, from Targ & Puthoff (1974)

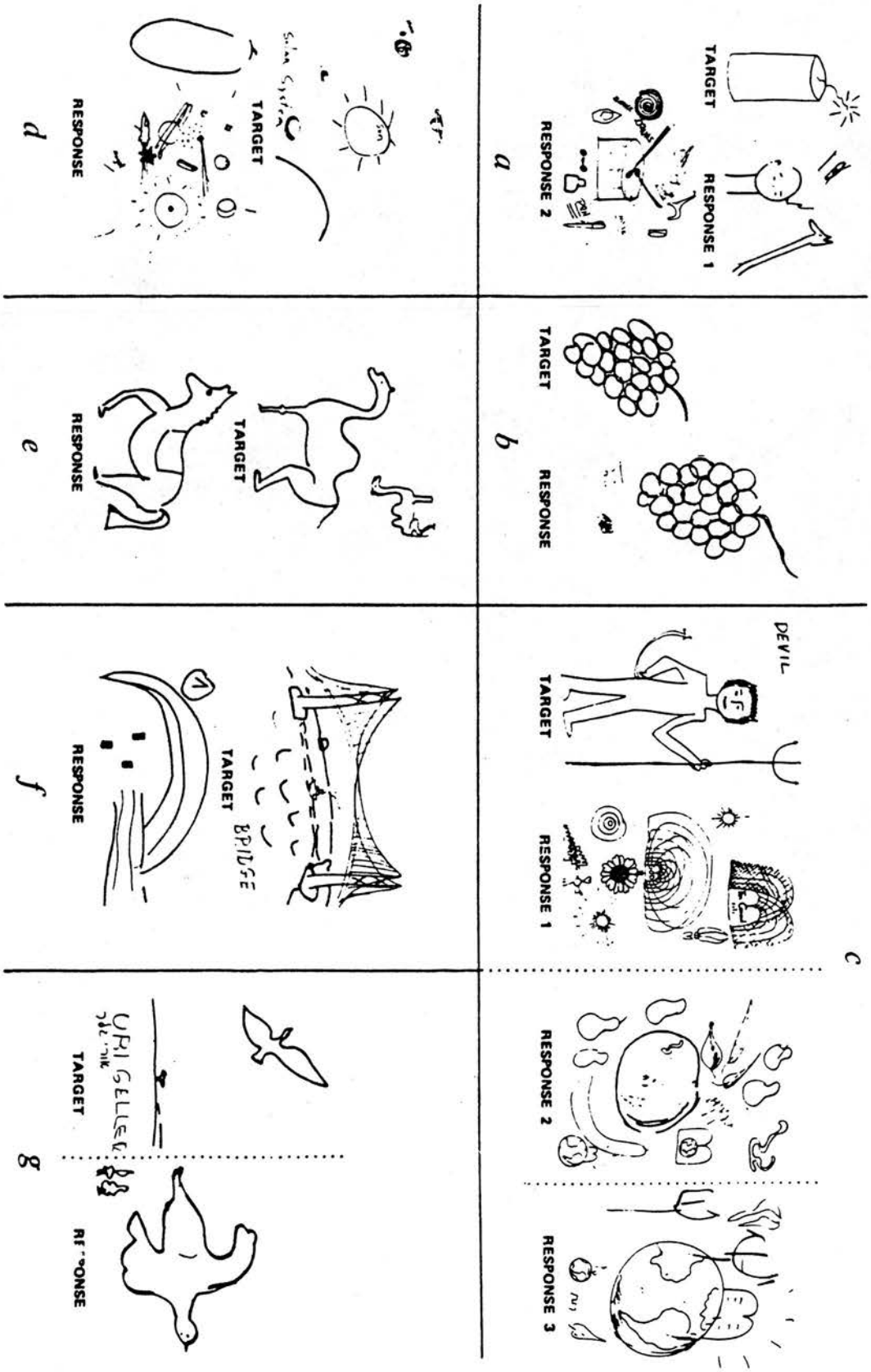


fig. 1

APPENDIX 8 (cont.)

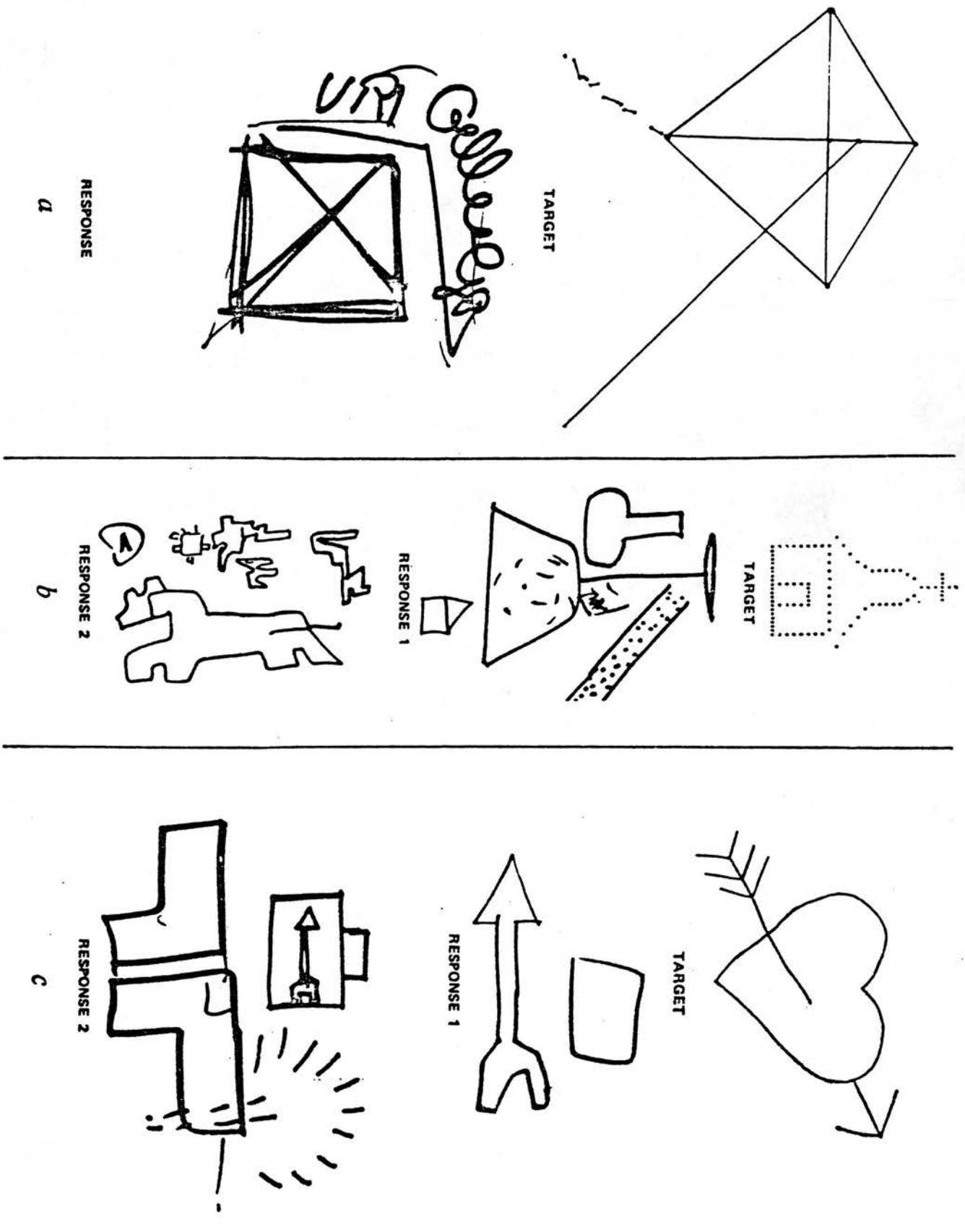


fig.2

APPENDIX 9

Transcript of tape-recorded instructions to judges.

"Welcome to the ESP experiment, Phase II. You're probably wondering what Phase I of the experiment was about, if you're in Phase II. Phase I was the actual telepathy test, in which I got a number of pairs of people to come in -- one pair at a time -- to participate in a picture-guessing test: one person in the pair was designated as the "sender" -- or "transmitter" -- and their task was to concentrate on a series of ten line-drawings -- three minutes each drawing; the "receiver" was in the next room: he or she had blank paper and a pen, and they had to attempt to reproduce the drawings being concentrated upon by the sender. This sounds like an impossible task; but if there really is such a thing as ESP -- the ability to perceive things without the use of the five ordinary senses -- then some small amount of knowledge should occasionally be transferred from the sender's mind to the receiver's mind, enabling the receiver to do drawings which, to some degree, resemble those of the sender. The big problem, in an experiment like this, is that we need some objective method of assessing the degree of similarity between the sender's drawing -- the "target" -- and the receiver's drawing -- the "response". For example, if the sender is concentrating upon a picture of a coin, and the receiver draws a circle, we want to know how likely it is that this is just a coincidence: if it is very unlikely, then it may be due to the receiver's obtaining a small amount of extra-sensory knowledge about the target-drawing, enabling him to draw the circle. So this is why we're running Phase II of the experiment: in Phase II, we get people -- like yourself -- to come in and act as a judge: you'll be faced with a whole set of targets and responses, and you'll have to try to match them up, on the basis of their similarity to each other; but more about that later.

"What I want to do first of all, is talk alittle bit about an experiment that was done at the Stanford Research Institute in California, with a man called Uri Geller. (Uri Geller, you might remember, is that man who seems to be able to bend metal -- spoons and forks, and so on.) Well, he also seems to be very good at guessing pictures in a situation like this. So what I'm going to do is start by showing you some of his data and talking alittle bit about it: I want to point out some important aspects which are necessary for you to know in this present experiment. If you look at the first one (Appendix 8, figure 1, a), you'll notice that the target was a fire-cracker. (The technical term, you'll remember, for the drawing that the sender concentrates upon is the "target", and whatever the receiver draws is called a "response".) I think Uri Geller -- who was of course the receiver in this case -- had any number of pieces of blank paper on which to draw his responses; our experiment was slightly different, inasmuch as the receiver had only one piece of blank paper, per target, but that's just a minor difference that we needn't worry too much about. If you look closely at Response Number 1, you'll see that what Uri Geller seems to have drawn is

something that looks like a little giraffe-neck, a little head on a neck with a curl, and what might be a hat flying up into the air. None of these things are an actual reproduction of the fire-cracker, per se. But if you look at the little head on the neck, with a curl (for example), topologically -- in other words, to do with the shape and the structure and dimensions of the drawing -- it's very similar to the fire-cracker, because you've got the two parallel lines, a circular or elliptoid piece up the top, and the little curl might even be the wick; you just don't have the curve at the bottom of the cracker to complete the drawing. The hat flying up into the air might be related to an explosion when the cracker goes off; and the giraffe-neck-like thing, at least visually -- in regard to its shape -- could, with some degree of imagination, be a match, to light the cracker in the first place.

"Well, we'll go on to Response Number 2. What he's drawn here are, amongst other things, a number of things that seem to be clearly wrong (for example, a little giraffe, and a pen, and so on); but his main feature is a drum, and once again, topologically, it's almost identical to the cracker: you only have to lengthen the two parallel sides alittle, and there you have the correct shape. And interestingly, he's written next to the drum the word "noise" -- which is exactly what the cracker makes when it goes off. So it seems that both in Response Number 1 and in Response Number 2, he's coming fairly close to both the shape of the target-drawing, and some of the concept at least, although he hasn't got either to a 100% degree accuracy.

"The next one we'll look at was the case where the target was the Devil holding the trident -- this is (figure 1) Response c. Uri Geller had three attempts at it altogether. If you look at Response Number 1, it looks as if he's drawn a number of irrelevant patterns, with no shape similarity at all to the target; but if you look closely, you'll notice the words "God" and "Ten Commandments" embedded in those patterns. So, what seems to have happened is that none of the visual information has come across in this case, but rather the concept or theme of religion. Now if that's so then we can understand Response Number 2, because what he's drawn here is, for example, an apple with a worm in it (right in the centre), something that looks like a serpent or a snake (right up in the top right-hand corner), various fruits and leaves that suggest the Garden of Eden, and small images of the world and light -- or perhaps the Ten Commandments -- growing out of it. It wasn't until Response Number 3 that he's actually done anything which is strikingly similar visually (and conceptually) to the target-drawing of the Devil, and that's the tridents of course; and in fact the whole drawing is dominated by his picture of the World with something bright -- the Ten Commandments, or something like that -- growing out of it. It's interesting that there are some little things right down the bottom of Response Number 3, which look like leaves, as suggested perhaps by his Response Number 2; but it's just possible that, in regard to their shape, they could be flames: now if that's so, then the symbolism of Hell is very interesting;

alternatively, they could be tongues of fire, which is a very important symbol in Christianity -- symbol of the Holy Spirit and the Paraclete, and so on. So in nearly every case, you can think of some sort of religious association in the things that are drawn in the responses, in relation to the target.

"So, by way of summary, what I want to point out about all this, is that there seem to be two major categories of response: on the one hand, what seems to come across is predominantly the shape features of the target-drawing: sometimes you can get all the shape reproduced, sometimes you can get just a part of the shape reproduced to a greater or lesser degree; alternatively, you can get all the shape reproduced plus a few things added on top of it -- embellished and added, things that weren't there in the original target-drawing. For example, the little head on a neck, with a curl, is an example where part of the target-drawing was missing, namely, the curve at the bottom of the cracker. And interestingly, what can happen when shape is transferred, is that what is drawn may look visually like the target-drawing, but the concept might be entirely wrong. For example, we've got the cracker, and the little head on the neck, again: the little head on the neck is very different conceptually from the fire-cracker: the drum is alot closer in that case because of the noise and the idea of noise that's similar between the two.

"So, just as an example of this sort of thing, we can have a look at the next drawing -- figure f -- where the target was a bridge: what he seems to have done in his response here, looks very similar to the carriage-way of the bridge, and perhaps the water underneath. This is an example of what we might call a "subtractive response", where part of the target-drawing is missing and he's only got little segments of it. Interestingly, also, this seems to be an example of where none of the concept has come across at all: I don't think Uri Geller had any idea that he'd drawn a bridge, or that the target-drawing was about anything like that: to him he's just drawn squiggles, and they just happen to visually resemble the target.

"So that's one major sort of response, where shape is the thing that predominates. The second major category of response is like that one where we talked about the Devil and religion and so on: what seems to happen here, is that the concept is transferred to the receiver's mind; when I say "the concept" I mean the ideas, the associations, the things that are conjured up in your mind when you contemplate the target-drawing -- things that are related to it on a conceptual level, an abstract level, rather than the concrete shape level. And just as in the shape example you can get ones where none of the concept comes across, in these concept relationship ones you can get none of the shape coming across either, so that you've got the two opposite extremes. Of course you can get combinations of the two: for example, if somebody does a perfect reproduction, then obviously they've got 100% of the shape and 100% of the concept. And in this regard we might point out to you his most famous response (this is Uri Geller again), and that's the bunch of

grapes -- figure b. This is his very best response ever: in the target-drawing there are apparently 24 grapes, and that's exactly the number of grapes that he's drawn in his response; so remarkable in fact that there's a great controversy over whether he might have cheated in some way, perhaps had a radio transmitter concealed in his tooth or something ridiculous like that. But anyway, that's a really striking response, because if you imagine if you were the receiver, of all the possible things that you could draw, it's very unlikely that you would do something as close to the target-drawing as what Uri Geller has done; if you come within cooee of it, -- if you do something which is at all similar, that's a very remarkable coincidence, statistically; and if you do something that's exactly right, well, the odds against that are simply astronomical.

"The next one we'll have a look at is figure e. This is an example of what I was saying about "hybrid" responses, where you can get some of the shape, and some of the concept too, to some degree: these two major categories of response are not mutually exclusive -- they overlap to some degree. And in this one, where the target was a camel, Uri Geller's done a horse: he's got very close to the shape, and, very close to the concept. It's abit difficult to say exactly what was happening in his mind to make him draw the horse: perhaps he got some sort of fuzzy mental image in his mind's eye of the camel, and misinterpreted it as a horse; or maybe, on the abstract, concept level, he got some idea that it was a four-footed beast of burden, and just decided that it was a horse. But it doesn't really matter, because in either case he's come very close to the target-drawing, and that's very good.

"The next one we'll have a look at is figure g. This one's interesting because, true enough, he has done one bird flying and one or two birds standing on the ground -- which is very very good -- but it's interesting to point out that probably what happened in this case is that he got the concept -- the idea -- rather than some sort of mental image of the target-drawing. And why I say that is just because the response-drawing is so different pictorially and visually, when you compare it with the target: it's not a good visual reproduction of the target: its more as if someone said to him "draw me one bird flying, and one or two birds standing on the ground", and that's what he's done. But nevertheless it's very interesting; he's come very close to it, and probably the question of what sort of information predominated in his mind is really of only theoretical importance: the fact is that he got it right. But I want to point out in all this the possibility of getting these concept responses as well as the shape ones: that's the important thing to note.

"The next one we'll look at -- figure d -- is a Solar System. And once again, he's got it perfectly right: he's drawn a solar system, although it's interesting that the details of what he's drawn are rather different to what is in the target-drawing. For example, his Saturn is rather different: the rings are different, he hasn't got any shadow on the planet's surface; his

APPENDIX 9 (cont.)

Sun is rather different inasmuch as the rays start some distance out from the circle, and he's got that funny black dot in the centre; and also he's done extra things that weren't present in the target-drawing: for example, a star, and a rocket, and shooting-stars, and things like that. So once again it's abit as if he got the idea, "draw a solar system", and that's what he's done, rather than getting some clear image of the target-drawing in his mind.

"Next one: we'll look at (figure 2), a. This is where the target-drawing was a kite. In its way, this one is almost as impressive as the grapes one, because what Uri Geller has drawn is very similar, shapewise, to the kite. It's not quite as impressive as the grapes one for two reasons -- two or three reasons in fact: first of all he hasn't got all of the shape: he's got the main structure of the kite but he perhaps hasn't got the string or the tail on it; secondly, it's not as complicated as the grapes one, so there aren't as many ways in which it could have gone wrong, and that's another reason why it's not as impressive, it's simpler than the target-drawing of the grapes; also, it's not quite clear whether Uri Geller was sure it was a kite: we're not sure whether he got the idea or not, because if you look at the response apart from the target, it just looks like a slightly skewed box with a cross in it: one can't be sure that he knew it was a kite; if he did, so much the better. And I suppose even so, he's got the shape pretty well right, so it's a very good response nevertheless.

"Next one we'll look at: Response Number c, where the target was an arrow through a heart. In his first response he has drawn an arrow, and he's drawn some sort of closed figure, although he hasn't got the idea that the closed figure is a heart, nor the idea that the arrow is through the heart. It's interesting that the arrow that he's drawn is very different visually from the one in the target-drawing: it's not as if he got some clear mental image (once again) of the target-drawing and that's what he drew; it's more as if he got the idea, "arrow", and that's his version of an arrow. Response Number 2: he's beginning to get the idea that the arrow's actually inside of, or on top of something, and what he's done here is to draw it on the outside of (I think it's supposed to be) a suitcase, or something like that. It's interesting that on the far right, these things that look like rays of the Sun or something, could possibly be fragments of visual information about the ends of the arrow: they look abit like the feathers on the end, although he hasn't got any idea that they're related to the arrow. The thing right down the bottom, on the left, -- the large feature -- is apparently a suitcoat, and if that's so, well, not very much you can say about that other than that it's not a very good match to the target-drawing; I suppose that you could say that it's symmetrical, like the heart is, and that it tapers towards the bottom, and your heart is underneath your coat when you're wearing it, but that's about all -- it's not as good a match as some of his others.

APPENDIX 9 (cont.)

"The final one we'll look at is figure b, where the target was the drawing of the church. It's in this dot formation because it was stored inside a computer, and that's what it looks like when it's put on the computer screen. It's interesting that in this case the computer-screen was turned off, and Uri Geller had to guess from out of the computer, just what the drawing was. And even in this very difficult case, he seems to have got rather close to it. For example, in Response Number 1 -- right down the bottom of Response Number 1 -- he's got a little house-like figure, and if it is supposed to be a house, well, both conceptually and shapewise, it's rather similar to the church. Similarly, on the far left, what he's done is something that is apparently a wine carafe, and wine and churches are related. On the far right you've got a whole array of dots: it looks as if he got some sort of visual information of these dots but didn't really know what to do with them. Right in the centre -- the large bit -- looks rather like the roof and the steeple of the church and perhaps most of the cross; and apparently the right way to hold it is upside down, so if you can invert it in your mind, it's apparently a wine-glass, or a chalice, and things like that are very related to altar-services, and churches, and so on.

"Response Number 2, down the bottom, is perhaps his poorest one: I think all the stuff on the left is irrelevant rubbish: you can't see any sort of similarity between that and the target-drawing. But the main feature, on the right, even this seems to have some sort of similarity: I don't think it's got any conceptual similarity, but it does seem to have elements of the shape represented: for example, the inset portion right at the bottom of the drawing could perhaps be the door; up the top of the drawing you've got a vertical line, and some oblique or horizontal line passing through it, and that could possibly be the cross; and even I suppose, the general shape -- the fact that it's almost rectangular down the bottom, and tapers up to a smaller section up the top, could possibly be what would be drawn if a very distorted, mangled image of the church came through. But so much for that one.

"The danger in going through and looking at all the targets and their responses in pairwise fashion like this, is that you can tend to read all sorts of similarities into the relationship that aren't really there -- they're abit tenuous and unjustified. So what we do in this present experiment -- Phase II -- is alot more careful than that. What I'll be doing is laying out all the ten responses, done by one particular receiver (remember there were ten target-drawings looked at by the sender, and ten response-drawings done by the receiver, at the same time as the sender was concentrating on the drawings). And what I'll be doing is giving you all the ten responses produced by one receiver, and giving you just one of the target-drawings (that's one of the drawings which the sender was concentrating upon at some time in the experiment.) The theory is, that if ESP has taken place in that particular trial, when the sender was concentrating on

that particular drawing, then the response that was done by the receiver should, to some extent, resemble it, even if it's only by a fairly small amount; but it should be sufficient that you, even though you don't know which is the right match, should be able to pick out the one that goes with that target. See what I mean? So, I'll lay out all the ten responses, give you one target, and ask you to pick out the one response which you think is most likely to have been drawn at the time when the sender was concentrating on that target. And bear in mind all the different sorts of similarity that we talked about in discussing Uri Geller's data: the shape similarities, the concept similarities, and the responses where you get a combination of the two. So, that will be your task today."

APPENDIX 10

PAIR : 5 TARGET-SET : "8"

RESPONSE / TARGET		I	II	III	IV	V	VI	VII	VIII	IX	X
		6403	7522	6319	2117	4707	6618	7669	4979	1931	4516
I	6435	5 7 2	7 4 5	1 2 1	6 9 7	4 8 8	2 6 3	10 10 10	3 5 6	8 3 9	9 1 4
		4.67									
II	6189	3 2 2	7 10 5	5 1 7	6 5 4	8 9 6	1 3 1	9 8 9	4 6 8	10 7 10	2 4 3
			7.33								
III	6370	9 2 9	4 3 5	8 9 4	7 8 3	1 1 1	10 7 6	5 10 8	3 5 2	2 4 10	6 6 7
				7.00							
IV	5925	6 1 1	2 9 6	3 4 5	1 3 4	10 8 10	4 5 2	7 2 9	9 10 7	8 7 8	5 6 3
					2.67						
V	4241	1	10	4	3	9	6	5	2	8	7
VI	8519										
VII	8185										
VIII	5272										
IX	3291										

APPENDIX 11

Pair: 62 Judge 1: Lúðvík Lárusson Judge 2: Halldór Bjarnason.

Trial	TARGET (13)	RESPONSE	\ SENDER	RECEIVER	JUDGE 1	JUDGE 2	Concordance
I	3560	8115	8641	8641	8641	8641	1.00
II	9855	4552	3672	3672	5261	8814	0.25
III	6770	9613	0702	8814	8814	6383	0.25
IV	6254	5261	5261 ✓	8115	9613	8115	0.25
V	4498	0107	6383	0107 ✓	0702	9613	0.00
VI	8310	0702	8814	9613	6383	5261	0.00
VII	3350	6383	8115	6383 ✓	8115	3672	0.25
VIII	6931	8641	9613	5261	3672	0702	0.25
IX	0945	3672	4552	4552	4552	4552	1.00
X	6554	8814	0107	0702	0107	0107	0.75
1		Σ correct	1	2	0	0	0.40

	S	R	J1	J2
S		3	4	3
R	3		3	3
J1	4	3		4
J2	3	3	4	

APPENDIX 12. AVERAGE RANKS : SHEEP (n=17)

	TRIAL											
PAIR	I	II	III	IV	V	VI	VII	VIII	IX	X	MEAN	S.D.
32	1.67	8.00	5.33	5.67	8.00	9.00	4.33	1.67	2.00	7.67	5.333	2.69
33	2.00	5.33	5.67	1.67	7.67	3.67	5.00	8.33	3.67	6.67	4.967	2.13
35	8.33	4.00	7.33	5.67	1.33	9.00	1.67	2.33	6.33	8.33	5.433	2.77
36	6.00	8.67	1.00	7.33	5.67	6.33	7.67	6.67	7.67	6.33	6.333	1.98
38	6.00	6.00	6.33	7.00	2.67	3.67	5.00	7.67	9.33	1.67	5.533	2.22
39	7.33	4.67	9.00	3.33	8.00	6.33	5.00	6.33	7.00	2.33	5.933	1.98
40	1.00	5.67	10.00	8.00	7.67	3.00	7.67	4.33	9.33	3.00	5.967	2.87
41	3.67	7.33	5.33	4.00	3.33	5.00	7.67	6.00	4.33	1.33	4.800	1.81
43	6.00	4.67	5.67	1.67	7.33	7.00	8.33	9.67	3.00	2.33	5.567	2.51
45	3.67	8.67	5.67	9.33	9.33	4.67	4.33	3.00	3.33	8.67	6.067	2.50
46	2.67	8.00	4.00	7.33	2.33	6.67	8.00	7.67	6.33	5.33	5.833	2.05
50	6.00	3.67	1.67	1.00	7.67	5.33	2.00	4.00	8.00	2.33	4.167	2.37
52	2.33	5.00	3.00	4.67	8.33	7.33	9.33	5.33	4.00	8.00	5.733	2.26
54	4.00	2.67	6.00	7.67	4.00	3.67	1.67	5.33	6.67	4.33	4.600	1.74
55	6.00	6.00	3.33	5.33	2.67	3.67	6.33	6.00	6.00	2.00	4.733	1.55
57	9.00	8.33	4.33	8.00	7.33	7.67	4.67	10.00	6.33	8.00	7.367	1.70
60	3.67	5.00	2.00	9.00	6.67	7.00	2.33	1.67	4.00	8.33	4.967	2.53
MEAN	4.667	5.980	5.039	5.686	5.882	5.824	5.353	5.647	5.726	5.098	5.490	0.76
S.D.	2.381	1.873	2.438	2.621	2.566	1.915	2.490	2.581	2.206	2.791	-	-
$\bar{X}$ hits	0.529	0.471	0.529	0.412	0.353	0.471	0.588	0.471	0.412	0.529	0.477	0.17
S.D.hits	0.515	0.515	0.515	0.507	0.493	0.515	0.507	0.515	0.507	0.515	-	-

APPENDIX 13. AVERAGE RANKS : GOATS (n=13)

	TRIAL											
PAIR	I	II	III	IV	V	VI	VII	VIII	IX	X	MEAN	S.D.
34	1.67	2.67	1.67	6.00	6.33	5.00	3.67	3.00	1.33	3.67	3.500	1.70
37	7.33	6.67	4.67	6.33	3.67	8.33	5.67	5.33	6.00	7.00	6.100	1.28
42	4.67	3.00	7.67	2.33	5.67	5.33	3.33	3.67	4.67	8.33	4.867	1.86
44	8.33	5.67	8.00	6.33	8.00	4.33	2.00	3.33	9.67	4.67	6.033	2.34
47	6.33	4.00	5.33	6.00	4.00	3.33	6.67	7.33	8.33	7.67	5.900	1.61
48	3.00	5.33	8.67	6.00	4.33	8.67	8.00	3.33	6.33	5.00	5.867	1.96
49	4.33	3.67	6.00	8.33	6.00	8.33	6.00	1.00	10.00	7.67	6.133	2.50
51	7.00	6.33	4.67	6.00	7.33	4.67	1.67	6.67	5.00	5.33	5.467	1.56
53	10.00	1.67	5.33	6.33	8.33	3.00	6.67	6.00	3.00	6.67	5.700	2.42
56	2.00	5.67	3.33	2.33	3.67	5.33	7.67	1.33	1.33	4.00	3.667	1.96
58	4.33	6.33	5.67	9.00	4.67	5.00	2.33	6.00	4.67	3.00	5.100	1.76
59	7.67	5.67	6.33	8.00	9.67	8.67	7.67	3.00	4.67	1.00	6.233	2.56
61	3.00	7.00	8.33	4.33	4.33	5.67	7.00	5.00	6.00	7.33	5.800	1.56
MEAN	5.359	4.897	5.821	5.949	5.846	5.821	5.256	4.231	5.452	5.487	5.413	0.91
S.D.	2.612	1.707	2.030	2.009	1.975	2.008	2.330	1.988	2.740	2.188	-	-
$\bar{X}$ hits	0.539	0.462	0.462	0.231	0.462	0.615	0.385	0.692	0.539	0.539	0.492	0.18
s.d.hits	0.519	0.519	0.519	0.439	0.519	0.506	0.506	0.480	0.519	0.519	-	-

APPENDIX 14. AVERAGE RANKS : NON-CLOSE-RELATERS (n=13)

	TRIAL											
PAIR	I	II	III	IV	V	VI	VII	VIII	IX	X	MEAN	S.D.
76	6.00	5.00	1.00	7.00	2.33	9.33	8.00	6.67	8.00	8.67	6.200	2.58
77	3.00	5.00	4.00	5.00	8.00	9.00	5.00	3.67	5.67	6.67	5.500	1.80
78	7.00	7.00	1.67	6.00	5.33	7.33	4.67	7.00	6.33	4.00	5.633	1.69
79	4.33	7.33	3.00	7.33	6.67	3.33	1.33	5.00	1.33	5.67	4.533	2.15
80	9.00	3.67	3.00	6.00	8.00	6.33	7.67	9.67	4.00	2.67	6.000	2.43
83	3.67	8.33	4.33	7.33	3.00	3.33	4.67	1.33	3.00	4.00	4.300	1.98
88	7.67	5.33	8.67	6.67	7.33	6.33	7.33	7.00	5.00	3.67	6.500	1.39
90	10.00	7.67	3.33	5.67	9.67	3.33	4.00	3.67	4.00	1.67	5.300	2.72
94	3.33	2.00	4.33	5.67	4.00	5.33	7.67	9.67	4.33	6.00	5.233	2.09
96	4.67	6.00	8.33	7.00	9.33	1.00	2.00	8.33	4.67	7.67	5.900	2.65
98	4.33	7.67	8.33	5.67	6.33	7.67	9.33	9.67	2.00	4.33	6.533	2.33
99	3.67	9.67	5.00	4.67	4.00	6.67	5.33	3.00	5.33	3.33	5.067	1.85
101	6.67	1.67	9.33	9.00	5.67	8.00	4.33	7.67	2.33	5.67	6.033	2.50
MEAN	5.641	5.872	4.949	6.385	6.128	5.923	5.487	6.333	4.308	4.923	5.595	0.70
S.D.	2.271	2.410	2.802	1.162	2.339	2.517	2.391	2.772	1.853	2.015	-	-
$\bar{X}$ hits	0.539	0.462	0.692	0.154	0.385	0.385	0.615	0.385	0.769	0.539	0.492	0.18
s.d.hits	0.519	0.519	0.480	0.376	0.506	0.506	0.506	0.506	0.439	0.519	-	-

APPENDIX 15. AVERAGE RANKS : STRANGERS (n=14)

	TRIAL											
PAIR	I	II	III	IV	V	VI	VII	VIII	IX	X	MEAN	S.D.
81	3.67	7.67	6.33	6.67	5.67	1.67	6.00	8.00	6.33	6.33	5.833	1.78
82	9.33	4.33	5.67	7.33	2.67	6.00	5.67	5.67	1.67	5.33	5.367	2.06
84	2.33	9.00	5.00	1.33	3.33	9.00	4.00	4.00	9.67	8.67	5.633	2.98
85	3.00	5.00	6.67	6.00	2.67	5.33	8.00	3.00	4.67	4.67	4.900	1.63
86	7.00	6.67	5.33	9.67	8.67	5.33	6.00	4.67	4.00	6.67	6.400	1.66
87	5.33	7.67	5.00	5.33	5.00	4.67	9.67	3.00	1.67	4.33	5.167	2.11
89	6.33	4.00	2.33	4.67	3.00	1.33	9.00	6.67	3.33	6.33	4.700	2.23
91	3.00	4.00	8.00	8.00	4.33	1.00	3.33	6.00	6.67	9.00	5.333	2.47
92	2.67	3.00	8.00	1.33	8.33	2.33	8.00	7.33	6.33	3.67	5.100	2.61
93	2.67	4.00	1.67	6.67	6.33	3.67	8.67	6.00	5.67	3.00	4.833	2.06
95	2.67	6.33	2.33	8.33	7.67	8.33	9.33	7.00	3.67	1.00	5.667	2.82
97	4.33	3.67	4.67	3.67	6.67	3.67	2.33	3.33	9.00	5.33	4.667	1.83
100	3.67	4.67	5.67	5.00	6.67	6.67	7.67	1.00	3.33	5.33	4.967	1.84
102	9.33	2.00	7.33	1.67	5.33	7.00	5.67	5.00	8.67	7.33	5.933	2.43
MEAN	4.667	5.143	5.286	5.405	5.452	4.714	6.667	5.048	5.333	5.500	5.321	0.52
S.D.	2.439	2.016	2.025	2.654	2.053	2.562	2.321	1.991	2.589	2.155	-	-
$\bar{X}$ hits	0.714	0.643	0.500	0.500	0.500	0.643	0.214	0.500	0.500	0.571	0.529	0.17
s.d.hits	0.469	0.497	0.519	0.519	0.519	0.497	0.426	0.519	0.519	0.514	-	-

APPENDIX 16. AVERAGE RANKS : MARRIED COUPLES (n=14)

	TRIAL											
PAIR	I	II	III	IV	V	VI	VII	VIII	IX	X	MEAN	S.D.
62	7.00	5.67	3.33	3.67	5.33	8.33	9.33	4.33	5.00	4.00	5.600	1.92
63	5.33	5.00	5.67	1.00	8.00	8.00	4.67	5.00	3.67	3.00	4.933	2.01
64	7.00	4.33	9.67	6.33	3.33	7.00	5.00	6.00	5.00	3.67	5.733	1.79
65	4.67	8.33	7.00	10.00	4.33	3.67	8.67	4.00	6.67	4.00	6.133	2.19
66	9.00	5.33	4.00	7.00	8.33	4.00	4.33	1.67	7.67	6.00	5.733	2.18
67	6.00	5.33	6.67	9.33	1.33	2.67	7.67	5.33	9.00	3.00	5.633	2.54
68	7.00	7.00	6.33	4.33	5.67	5.00	2.33	3.67	9.00	8.67	5.900	2.02
69	9.67	7.33	3.67	5.67	5.00	5.00	3.00	5.33	3.67	4.33	5.267	1.87
70	5.00	5.67	7.33	3.00	8.67	7.00	1.33	7.33	7.00	9.00	6.133	2.31
71	3.33	2.33	3.00	5.00	8.67	5.67	5.00	3.33	8.00	5.00	4.933	1.99
72	6.67	10.00	5.00	8.33	3.33	7.33	7.67	3.33	7.00	3.00	6.167	2.27
73	7.67	8.33	5.67	8.33	6.33	4.00	2.00	7.00	2.67	3.00	5.500	2.29
74	1.33	4.67	8.33	5.67	2.67	7.33	9.33	1.67	5.00	3.33	4.933	2.62
75	7.00	8.67	9.00	6.33	10.00	3.00	3.33	5.67	2.00	4.33	5.933	2.61
MEAN	6.191	6.286	6.048	6.000	5.786	5.571	5.262	4.548	5.810	4.595	5.610	0.45
S.D.	2.171	2.071	2.112	2.522	2.640	1.923	2.789	1.738	2.282	1.992	-	-
$\bar{X}$ hits	0.357	0.429	0.357	0.357	0.500	0.500	0.643	0.714	0.500	0.786	0.514	0.13
s.d.hits	0.497	0.514	0.497	0.497	0.519	0.519	0.497	0.469	0.519	0.426	-	-

APPENDIX 17. AVERAGE TARGET-RANKS : CLOSE-RELATERS (n=18)

	TRIAL											
PAIR	I	II	III	IV	V	VI	VII	VIII	IX	X	MEAN	S.D.
1	7.00	1.67	4.33	7.67	9.33	3.67	7.00	8.00	7.33	7.33	6.333	2.22
3	2.00	6.00	5.33	5.00	8.00	3.67	6.33	7.00	7.00	2.00	5.233	1.98
4	5.67	5.00	7.00	2.00	7.67	2.00	4.67	4.67	4.33	6.67	4.967	1.82
5	4.33	7.67	5.00	5.00	1.67	5.00	4.33	5.33	8.67	4.00	5.100	1.83
7	7.00	6.00	4.67	1.33	3.00	4.33	8.33	9.67	6.00	4.00	5.433	2.37
9	7.67	6.00	8.00	7.67	3.67	9.00	8.33	5.00	5.33	9.33	7.000	1.79
10	8.00	2.33	3.00	4.00	1.33	9.67	2.33	4.33	2.33	7.00	4.433	2.68
11	8.00	8.67	8.67	7.67	1.00	10.00	10.00	5.67	6.33	8.67	7.467	2.53
12	7.33	7.67	9.00	9.33	1.00	9.33	3.67	5.67	7.33	2.67	6.300	2.80
13	3.67	7.00	4.00	7.00	1.33	7.00	1.67	4.33	7.00	5.67	4.867	2.10
14	7.33	3.33	4.00	2.33	4.67	4.67	2.33	10.00	2.67	5.33	4.667	2.30
16	7.00	6.00	3.00	4.00	1.33	4.67	4.00	9.00	8.67	6.67	5.433	2.34
17	6.67	2.00	9.00	8.00	9.67	4.67	5.00	3.00	6.00	4.67	5.867	2.37
20	4.33	4.33	1.00	6.33	8.00	9.33	5.00	9.00	7.67	9.33	6.433	2.60
22	5.67	4.67	10.00	5.33	5.00	7.33	5.33	8.33	4.67	4.33	6.067	1.78
25	6.67	6.67	9.33	7.00	6.00	3.00	6.67	2.67	8.67	3.33	6.000	2.19
30	3.67	4.33	1.00	5.33	4.33	4.33	6.00	1.00	5.00	8.00	4.300	2.01
31	5.00	7.00	2.67	5.67	1.33	2.67	7.00	6.00	6.33	2.00	4.567	2.07
MEAN	5.944	5.352	5.500	5.593	4.352	5.796	5.444	6.037	6.185	5.611	5.582	0.91
S.D.	1.746	2.047	2.929	2.233	3.090	2.674	2.249	2.567	1.876	2.396	-	-
$\bar{X}$ hits	0.333	0.444	0.611	0.500	0.667	0.611	0.556	0.444	0.333	0.500	0.500	0.20
s.d.hits	0.485	0.511	0.502	0.515	0.485	0.502	0.511	0.511	0.485	0.515	-	-

APPENDIX 18. AVERAGE TARGET-RANKS: NON-CLOSE-RELATERS (n=13); SIGNIFICANT PAIRS FROM FIRST SHEEP-GOAT STUDY (n=5)

	TRIAL											
PAIR	I	II	III	IV	V	VI	VII	VIII	IX	X	MEAN	S.D.
2	3.67	6.00	7.67	7.67	7.67	8.00	4.67	6.33	5.00	9.33	6.600	1.68
6	7.67	3.67	5.33	9.67	7.33	6.67	8.00	5.33	7.00	7.00	6.767	1.58
8	2.00	4.67	7.00	5.00	3.00	3.33	5.67	3.00	2.33	1.00	3.700	1.75
15	7.67	3.33	5.33	9.33	3.00	5.33	6.00	7.67	3.33	4.33	5.533	2.03
18	4.67	8.00	4.67	7.33	2.33	7.00	8.67	8.00	9.67	9.33	6.967	2.24
19	4.33	3.33	7.00	7.33	8.67	9.33	3.67	4.67	4.00	6.33	5.867	2.05
21	6.00	5.67	7.00	6.00	8.33	5.00	2.33	4.33	8.00	2.33	5.500	1.97
23	5.67	5.00	5.00	8.00	9.00	8.33	2.00	5.33	2.33	4.67	5.533	2.24
24	9.33	6.00	8.00	7.67	6.33	6.00	3.33	8.00	8.33	5.33	6.833	1.67
26	3.00	7.67	8.67	8.00	7.67	3.67	6.00	4.00	1.00	5.67	5.533	2.41
27	4.00	6.67	7.00	4.00	8.00	2.67	1.67	3.33	7.67	2.67	4.767	2.22
28	6.67	5.33	3.67	5.67	8.67	4.33	7.33	4.00	4.67	9.00	5.933	1.81
29	3.67	5.67	2.67	5.33	2.33	7.67	4.67	3.33	8.67	10.00	5.400	2.48
MEAN	5.256	5.462	6.077	7.000	6.333	5.949	4.923	5.180	5.539	5.923	5.764	0.91
S.D.	2.131	1.488	1.780	1.689	2.639	2.094	2.278	1.798	2.840	2.929	-	-
$\bar{X}$ hits	0.539	0.462	0.462	0.231	0.308	0.462	0.539	0.692	0.539	0.462	0.469	0.17
s.d.hits	0.519	0.519	0.519	0.439	0.480	0.519	0.519	0.480	0.519	0.519	-	-
34	5.67	4.00	6.33	2.33	7.00	6.00	8.33	5.67	1.00	3.33	4.966	2.14
36	8.33	7.00	1.33	6.67	9.67	3.00	8.00	7.33	9.00	2.33	6.266	2.80
40	1.33	7.00	6.00	7.67	2.00	8.00	7.33	5.00	4.00	3.33	5.166	2.29
56	1.00	3.33	1.33	3.67	4.67	8.33	7.67	1.00	7.33	1.33	3.966	2.77
57	9.00	6.67	3.00	7.33	7.67	9.00	5.33	6.33	7.33	7.67	6.933	1.69

Appendix 19. AVERAGE RANK-SCORES FOR FOUR EXPERIMENTS SUCCESSIVELY COMBINED (n=102)

EXPERIMENT	TRIAL I	II	III	IV	V	VI	VII	VIII	IX	X	MEAN
1st (n=31)	4.710	5.032	5.828	5.452	5.742	6.828	5.581	5.333	5.559	6.011	5.608
2nd (n=30)	4.967	5.511	5.378	5.800	5.867	5.822	5.311	5.033	5.611	5.267	5.457
1+2 (n=61)	4.836	5.269	5.607	5.623	5.803	6.333	5.448	5.186	5.584	5.644	5.533
3rd (n=27)	5.136	5.494	5.124	5.877	5.778	5.296	6.099	5.667	4.840	5.222	5.453
1+2+3(n=88)	4.928	5.338	5.458	5.701	5.796	6.015	5.648	5.334	5.356	5.515	5.509
4th (n=14)	6.191	6.286	6.048	6.000	5.786	5.571	5.262	4.548	5.810	4.595	5.610
1+2+3+4 (n = 102)	5.101	5.468	5.539	5.742	5.794	5.954	5.595	5.226	5.418	5.389	5.523

Appendix 20. AVERAGE HIT-SCORES FOR FOUR EXPERIMENTS SUCCESSIVELY COMBINED (n=102)

EXPERIMENT	TRIAL I	II	III	IV	V	VI	VII	VIII	IX	X	MEAN
1st (n=31)	0.613	0.548	0.355	0.581	0.387	0.161	0.548	0.516	0.516	0.452	0.468
2nd (n=30)	0.533	0.467	0.500	0.333	0.400	0.533	0.500	0.567	0.467	0.533	0.483
1+2 (n=61)	0.574	0.508	0.426	0.459	0.393	0.344	0.525	0.541	0.492	0.492	0.475
3rd (n=27)	0.630	0.556	0.593	0.333	0.444	0.519	0.407	0.444	0.630	0.556	0.511
1+2+3(n=88)	0.591	0.523	0.477	0.420	0.409	0.398	0.489	0.511	0.534	0.511	0.486
4th (n=14)	0.357	0.429	0.357	0.357	0.500	0.500	0.643	0.714	0.500	0.786	0.514
1+2+3+4 (n = 102)	0.559	0.510	0.461	0.412	0.422	0.412	0.510	0.539	0.529	0.549	0.490

Key:
'1st' = The Closeness Study
'2nd' = The First Sheep-Goat Study
'3rd' = The Second Sheep-Goat Study
'4th' = The Married Couple Study

APPENDIX 21.

Published reports, in chronological order, of experiments which in whole or in part have employed the drawing-reproduction technique:

(1) Gurney, Myers and Barrett (1883)
(2) Gurney, Myers, Podmore and Barrett (1883)
(3) Gurney, Myers and Barrett (1884a)
(4) Gurney, Myers and Barrett (1884b)
(5) Guthrie (1884)
(6) Lodge (1884a)
(7) Lodge (1884b)
(8) Pickering (1885)
(9) Guthrie (1885)
(10) Dessoir (1886)
(11) Schmoll (1887)
(12) Schmoll & Mabire (1888)
(13) Richet (1888)
(14) Dessoir (1888)
(15) Schrenck-Notzing (1891)
(16) Lodge (1892)
(17) Rawson (1895)
(18) Usher & Burt (1909)
(19) Warcollier (1921)
(20) Dingwall (1924)
(21) Tischner (1925)
(22) Brück (1925)
(23) Sinclair (1930)
(24) Prince (1932)
(25) Ossowiecki (1933)
(26) Besterman (1933)
(27) Pratt (1937)
(28) Warcollier (1938)
(29) Carington (1940a)
(30) Carington (1940b)
(31) Carington (1941)
(32) Stuart (1942)
(33) Stuart (1944)
(34) Carington (1944a)

APPENDIX 21 (continued)

(35) Carington (1944b)
(36) Carington & Heywood (1944)
(37) Stuart (1945a)
(38) Stuart (1945b)
(39) Taves, Murphy and Dale (1945)
(40) Stuart (1946)
(41) Smith & Humphrey (1946)
(42) Humphrey (1946a)
(43) Humphrey (1946b)
(44) McMahan (1946)
(45) Stuart (1947)
(46) Stuart, Humphrey, Smith and McMahan (1947)
(47) Nash & Richards (1947)
(48) Bevan (1947a)
(49) Bevan (1947b)
(50) West (1947)
(51) Warcollier (1948)
(52) Schmeidler & Allison (1948)
(53) Rush & Jensen (1949)
(54) West (1950)
(55) Cadoret (1953)
(56) Marsh (1962)
(57) Borzymowski (1962)
(58) Schwarz (1963)
(59) Borzymowski (1965)
(60) Pratt (1966)
(61) Beloff & Mandleberg (1967)
(62) Barry (1971)
(63) Hardy, Harvie and Koestler (1973)
(64) Musso & Granero (1973)
(65) Targ & Puthoff (1974)
(66) Vellissaris & Vellissaris (1977)
(67) Mittenecker & Schulter (1978)
(68) Braud, Davis and Wood (1979)
(69) Thalbourne (1979a)
(70) Thalbourne (1980)
(71) Hasted (1981)

APPENDIX 22: Program RANMAT

```
C       PROGRAM RANMAT(INPUT,TAPE1=INPUT,OUTPUT)
        DIMENSION AVJ(10,10),INR(30),ISEL(10,10),IC(10),IR(10),
     1  XMN(1000),XSD(1000)
        TYPE 999
999     FORMAT('ØDATA INPUT FILE:  ')
        CALL ASSIGN(1,,-1)
        TYPE 888
888     FORMAT(' DO YOU WANT ENTIRE FILE RUN?')
        TYPE 889
889     FORMAT(' YES  (TYPE 1)',/,' NO (TYPE 0)',/,'Ø* ')
        ACCEPT 890,KTYPE
890     FORMAT(I2)
        IF(KTYPE.EQ.1)GO TO 891
        TYPE 893
893     FORMAT('ØPROCESSING STARTS WITH SEQUENCE NO.? * ')
        ACCEPT 890,ILOOP
        ILOOP=ILOOP-1
        DO 894 N=1,ILOOP
        DO 897 I=1,10
        READ(1,10)ISUB,INR
897     CONTINUE
894     CONTINUE
891     N=10
        NM1=N-1
        NP1=N+1
        FN=N
        JRAN=0
        KRAN=0
        T=RAN(JRAN,KRAN)
5       SX=0.0
        SXX=0.0
        DO 40 I=1,N
        READ (1,10,END=120)ISUB,INR
10      FORMAT(A2,2X,30I2)
20              DO 30 J=1,N
                K=J*3-2
                AVJ(I,J)=(INR(K)+INR(K+1)+INR(K+2))/3.0
C               AVJ(I,J)=ABS((INR(K)+INR(K+1)+INR(K+2))/3.0 -5.5)
30              ISEL(I,J)=0
        X=AVJ(I,I)
        SX=SX+X
40      SXX=SXX+X*X
        DMN=SX/FN
        DSD=SQRT((SXX-SX*SX/FN)/FN)
        DO 50 I=1,N
        IR(I)=I
50      IC(I)=I
        DO 70 I=1,1000
        SX=0.0
        SXX=0.0
```

APPENDIX 22 (continued)

```
                DO 60 J=1,NM1
                L=NP1-J
                K=INT(L*RAN(JRAN,KRAN))+1
                II=IR(K)
                IR(K)=IR(L)
                IR(L)=II
                K=INT(L*RAN(JRAN,KRAN))+1
                JJ=IC(K)
                IC(K)=IC(L)
                IC(L)=JJ
                X=AVJ(II,JJ)
                SX=SX+X
                SXX=SXX+X*X
60              ISEL(II,JJ)=ISEL(II,JJ)+1
        II=IR(1)
        JJ=IC(1)
        X=AVJ(II,JJ)
        SX=SX+X
        SXX=SXX+X*X
        ISEL(II,JJ)=ISEL(II,JJ)+1
        XMN(I)=SX/FN
70      XSD(I)=SQRT((SXX-SX*SX/FN)/FN)
        TYPE 80,ISUB
 80     FORMAT(1H1,///,10X,'SUBJECT  ',A4,//,10X,'AVERAGE JUDGEMENT
C 80    FORMAT(1H1,///,10X,'SUBJECT  ',A4,//,10X,'ABSOLUTE DEVIATION
     1AND CELL SELECTION MATRIX')
        TYPE 90,(I,I=1,N),(I,(AVJ(I,J),J=1,N),(ISEL(I,J),J=1,N),
     1  I=1,N)
90      FORMAT(10X,10I8/(/,I10,10F8.2/10X,10I8))
        TYPE 100,ISUB
100     FORMAT(1H1,///,10X,'SUBJECT  ',A4,//10X,
     1  'RANDOMIZATION STATS FOR THE MEAN')
        CALL RANSTA(DMN,1000,XMN,.5,1.0,20.0)
C       CALL RANSTA(DMN,1000,XMN,.5,0.0,20.0)
        TYPE 110
110     FORMAT(//10X,'RANDOMIZATION STATS FOR THE S.D.')
        CALL RANSTA(DSD,1000,XSD,.1,0.0,45.0)
        GO TO 5
120     STOP
        END
C
C
        SUBROUTINE RANSTA(DX,N,X,DIV,ZR,ZMX)
        DIMENSION X(N),IH(100),D(7),DEC(7)
        LOGICAL SWAP
        DATA D/.01,.025,.05,.50,.95,.975,.99/
        FN=N
        K=N
10      SWAP=.FALSE.
        K=K-1
        DO 20 I=1,K
        J=I+1
        IF(X(I).LT.X(J)) GO TO 20
```

APPENDIX 22 (continued)

```
          T=X(I)
          X(I)=X(J)
          X(J)=T
          SWAP=.TRUE.
20        CONTINUE
          IF(SWAP.AND.K.GT.1) GO TO 10
          DO 30 I=1,100
30        IH(I)=0
          SX=0.0
          SXX=0.0
          IE=0
          IG=0
          DO 40 I=1,N
          T=X(I)
          SX=SX+T
          SXX=SXX+T*T
          IF(DX.GT.T) IG=IG+1
          IF(DX.EQ.T) IE=IE+1
          J=MIN1((T-ZR)/DIV+1,ZMX)
40        IH(J)=IH(J)+1
          DO 50 I=1,7
          J=INT(N*D(I))+1
50        DEC(I)=X(J)
          AM=SX/N
          SD=SQRT((SXX-SX*SX/N)/N)
          MAX=ZMX
          XG=IG/FN
          XE=IE/FN
          TYPE 60,DX,AM,SD,XE,XG,DEC,(IH(I),I=1,MAX)
60        FORMAT(/5X,'DIAGONAL STATISTIC',F6.2,
     1    /5X,'DISTRIBUTION MEAN AND S.D.',2F6.2,
     2    //5X,'PROPORTION EQUAL',F7.3,
     3    /5X,'PROPORTION LESS',F7.3,
     4    //5X,'CENTILE POINTS',7F6.2,
     5    //5X,'GROUPED FREQUENCY DATA'/(5X,20I5))
          RETURN
          END
```

Miscellaneous notes:

```
AVJ  :  average judgement
INR  :  number of rank-scores
ISEL :  counter for number of times randomly selected
IC   :  number of columns
IR   :  number of rows
XMN  :  number of randomization-test means
XSD  :  number of randomization-test s.d.'s
N    :  number of records
FN   :  number of average-ranks on principal diagonal
X    :  average-rank on principal diagonal
SX   :  sum of average-ranks on principal diagonal
SXX  :  sum of squared average-ranks on diagonal
DMN  :  mean of average-ranks on principal diagonal
DSD  :  s.d. of average-ranks on principal diagonal
```

APPENDIX 23.

Review of all experiments to date (outside this thesis) in which the Sheep-Goat Scale has been used.

Harding (1980); Harding & Thalbourne (1981)

Susan Harding, of the University of Edinburgh, conducted an experiment to investigate whether the practice of the technique known as Transcendental Meditation (TM) was associated with higher scores on both forced-choice and free-response clairvoyance tests. The two ESP tasks consisted of two runs of Zener cards, and a single attempt to identify and draw the contents of an enclosed coloured target-picture; these tests were administered in counter-balanced design both before and after 20 minutes of either rest or meditation. There were, in all, 24 subjects: ten control subjects (who were required simply to rest for 20 minutes), nine ordinary Transcendental Meditators (who were required to meditate for 20 minutes), and five so-called 'sidhas' -- persons who had received training in a supposedly advanced form of TM known as the TM-Sidhi program -- (who were asked to meditate and perform the sidhi for obtaining knowledge of hidden objects). The eleven-item Sheep-Goat Scale was administered prior to the ESP tests.

Since no before-after differences in ESP-scoring were observed, we shall restrict the discussion to the average card-guessing and picture-guessing scores obtained on the two testing-occasions. No evidence of ESP was found at any stage of the experiment, either overall, or between the three subject-groups, or when the 24 Sheep-Goat Scale-scores and the component items were correlated with the ESP-scores; the only pattern to emerge from the sheep-goat analysis was that the correlations tended to be positive with the free-response scores but negative with the forced-choice scores (despite the low positive correlation between card-scores and picture-guessing scores).

However, one might conjecture that the SGE could have occurred in some subject-groups but not in others; accordingly, the correlations were re-performed for control subjects, meditators and sidhas separately. But again, no extra-chance number of significances was obtained; the only observation of interest is that the correlations tended to be positive for meditators, and negative both for sidhas and for controls, these trends being

consistent across card-guessing and picture-guessing. Thus, a post hoc speculation for the future might be that the most promising group for showing the SGE might be ordinary meditators, especially if a free-response test is used. But the blunt fact of the matter is that in this context at least, the Sheep-Goat Scale was not able to predict clairvoyance-scores to a significant degree.

Delanoy (1981)

Deborah Delanoy, also of the University of Edinburgh, attempted to bring about an incline in the GESP-scoring of six subjects by exposing them repeatedly (viz., on 12 occasions) to the so-called 'ganzfeld' technique, in which homogeneous sensory stimulation is used in the hope of inducing a hypnagogic-like state, thereby rendering subjects more aware of internal imagery (some of which may be extrasensory in origin). The subjects were specially selected by virtue of their being 'extraverted sheep' (i.e., they scored above the student population average both on Eysenck's Extraversion-scale (Eysenck & Eysenck, 1964), and on the (ten-item version of the) present author's Sheep-Goat Scale); they were thus thought to possess personality characteristics which are at least slightly associated with psi-hitting. The ESP test was a free-response one, using pictorial targets, and with the experimenter acting as agent; detailed feedback was provided to the subject after each trial so as to render them more aware of inappropriate introspection-strategies. The 12 trials of each subject tended to be spread out over a number of weeks.

No overall psi-hitting was observed for the group as a whole, nor was there any evidence of scoring-inclines. Thus, neither the Sheep-Goat Scale nor the Extraversion-Scale was able to pre-select a group which would psi-hit. Nevertheless, there was, post hoc, a significantly positive correlation between ESP-score and extraversion-score, despite the highly restricted range of the latter: $r = +.74$, $n = 6$, $1p = 0.047$; indeed, the most extraverted subject psi-hit, and the most introverted member (even though he was slightly above average in extraversion compared to the student population) psi-missed. The Sheep-Goat Scale-scores, however, showed a non-significant negative correlation with ESP-score ($r = -.46$). A possible reason for the failure of the Sheep-Goat Scale in this context is that agent and percipient rapidly became

good friends in the course of the 12 ganzfeld sessions -- a social situation which we have seen is inimical to the occurrence of the SGE. Using a barely-acquainted stranger as agent, preferably a sheep, might have been more likely to result in the desired above-chance scoring.

Makris (1979)

Mary Makris, of the University of Adelaide, conducted a so-called 'remote viewing' experiment, the percipients' task being to describe the surroundings of a distant agent/experimenter team. Prior to this test of GESP, all percipients (but not, regrettably, the agents) were administered a twelve-item version of the Sheep-Goat Scale (there being two questions on experience of telepathy, which were combined in a logical fashion to form just one item so as to make the Scale more comparable to the standard eleven-item one.) Each of the ten consecutive trials lasted approximately 50 minutes long (sic), and there was feedback to the percipients after every trial. For each of the 16 percipients, the person who acted as his or her agent was in every case a stranger: eight of the percipients were 'close-relaters' (i.e., persons who reported having a close, mutually reciprocated relationship with someone (though not, of course, with the agent)); eight percipients were non-close-relaters.

The object of the experiment was to observe whether 'being a close-relater' was associated with high GESP-scores even when working with a stranger as agent. This hypothesis was confirmed: close-relater percipients scored significantly above chance, and significantly higher than the non-close-relaters. It thus seems plausible to suggest that since close-relaters can show evidence of psi-hitting even with a stranger, the supposed frequency of psi-hitting within a mutually close dyad may be due not to the existence of a close relationship so much as the 'outgoing' personality-types of the sort of people who tend to form such relationships. However, it should be pointed out that the lustre of these significant results is tarnished somewhat by the presence of at least one potentially controversial avenue of sensory leakage which might have led to spurious confirmation of the experimenter's hypotheses. Nevertheless, a hugely significant serial-position effect was also found; quite contrary to Makris' expectations,

this turned out to be an inverted U-shape; this fact makes it seem rather more unlikely that sensory leakage could account for this effect.

Furthermore, Makris did not even think to look for evidence of a sheep-goat effect; she did, however, make her data available to the present author for analysis. It was promising that for all 16 subjects, the agent-percipient relationship was of just the sort postulated by the author to be a sine qua non for the occurrence of the SGE in a free-response GESP test. As it turned out, the correlation was in the right direction but not significant: for mean rank, r = -.32, for hits, +.26. It was additionally supposed that the size of the sheep-goat/ESP correlation would be comparable for the two sub-groups -- close- and non-close-relaters -- since in both the agent was a barely acquainted stranger; yet post hoc, this was found not to be the case: for the eight non-close-relaters, the correlations were non-significantly negative (r = +.02 for mean rank, -.08 for hits), but for the eight close-relaters they were comfortably significant (r = -.92 for mean rank (2p = 0.001), and r = +.85 for the hits measure (2p = 0.008)). Thus, there was at least partial confirmation of the sheep-goat effect. Why the two ostensibly equivalent subgroups should have yielded such discrepant results, is for the time being utterly anomalous; nevertheless, this experiment suggests that even if the absence of a close relationship between sender and receiver is necessary for the SGE, of itself it is not sufficient for its elicitation.

Thalbourne, Beloff and Delanoy (1982)

The basic premise in this study was that psi-prediction may perhaps be enhanced by combining psychological variables known on independent grounds to be (1) capable of predicting psi, and (2) related to each other empirically. Now there have in fact been observed small but significant positive correlations between extraversion and the sheep-goat variable (Thalbourne & Haraldsson, 1980; Thalbourne, 1981); at the same time, the Sheep-Goat Scale used in these personality studies seems to be able under certain circumstances to predict GESP-scoring, while Eysenck (1967) provides some evidence that his Extraversion-Scale may be predictive of psi (cf. Delanoy, 1981, above).

Thus, two samples of psychology students were administered the (ten-item) Sheep-Goat Scale, Eysenck's Extraversion-Scale, and a ten-trial forced-choice test of precognition called 'Consumer's Choice'. Both psychological measures were dichotomized around their population means so as to divide subjects into sheep and goats, and into extraverts and introverts. The major prediction was that the 25 sheep who were also extraverted would score significantly higher than the 17 goats who were introverted, these two subgroups also scoring significantly positively and negatively respectively.

In the first sample (n = 101), marginal significance was attained for the difference between extraverted sheep and introverted goats (1p = 0.058), and significance for the below-chance mean of the latter group (1p = 0.005). Post hoc, significant negative scoring was found for the entire group (n = 101, 2p = 0.035), for the 31 introverts (2p = 0.022), and for the 62 goats (2p = 0.033). Surprisingly, the Extraversion- and Sheep-Goat Scales were negatively correlated with each other; the correlation between Extraversion and ESP was negative, and between Sheep-Goat and ESP positive; none of these three correlations, however, was significant.

In the second sample (n = 86), extraverted sheep again scored higher than introverted goats, but the difference was non-significant, and neither did the two means differ significantly from chance. Even so, the 32 sheep scored significantly higher than the 54 goats (2p = 0.027), with the latter group scoring significantly negatively (1p = 0.045). Post hoc, there was found to be a significantly positive correlation between precognition-scores and scores on the (ten-item) Sheep-Goat Scale: r = +.30, 2p = 0.005). In this second study, an eleventh sheep-goat item was administered (concerning the subject's estimate of the likelihood of they themselves demonstrating ESP in the test); this item is normally included in the Sheep-Goat Scale when the Scale is being administered just prior to an ESP test: by oversight it had been omitted in the case of the first sample; it was included for this second sample, but phrased retrospectively and administered after completion of the precognition test; its addition to the ten-item Scale actually reduced the significance of the correlation (r = +.24, 2p = 0.027), in contrast to its usual effect of enhancing it when administered prospectively (i.e., prior to the ESP test.)

A similar reduction in significance of the correlation with ESP was observed upon adding four other exploratory sheep-goat items (most of these concerning abstract belief rather than experience of various types of cognitive psi). Nevertheless, it may be of interest to note that all of the ten core sheep-goat items correlated positively with ESP-score, three of them significantly so, namely, personal experience of ESP, experience of telepathy, and the claim to be psychic (cf. Table 7.1).

On the one previous occasion when the Sheep-Goat Scale was used in a non-GESP non-free-response context (namely, Harding, 1980), no significant correlations were found; it is thus all the more surprising, and yet encouraging, that new ground should be broken here by achieving success in the prediction of forced-choice precognition scores. Thankfully, this study demonstrates that the psi-predictive ability of the Sheep-Goat Scale is not restricted to tests of general extrasensory perception of the contemporaneous mode.

R E F E R E N C E S

AKOLKAR, V.V. & DESHPANDE, V.D. (1966) Paranormal perceptivity and ESP test performance. Journal of Parapsychology, Vol.30, no.2, p.131

BARRINGTON, M.R. (1973) A free response sheep/goat experiment using an irrelevant task. Journal of the Society for Psychical Research, Vol.47, no.758, pp.222-245

BARRY, J. (1971) Journal d'un Parapsychologue (Diary of a Parapsychologist.) Editions et Publications Premieres, Evreux, France. Reviewed by K.M. Banham in Journal of Parapsychology, 1973, Vol. 37, no.1, pp.55-56

BEER, D. (1971) A correlational study to determine the effects of marital status on 'telepathy' (GESP) between individuals. Journal of Parapsychology, Vol.35, no.2, p.157

BELOFF, J. (1969) The 'sweethearts' experiment. Journal of the Society for Psychical Research, Vol.45, no.739, pp.1-7

BELOFF, J. (1973a) Belief and doubt. Presidential address to the 15th Annual Convention of the Parapsychological Association. In W.G. Roll, R.L. Morris and J.D. Morris (Eds.), Research in Parapsychology 1972. Scarecrow Press, Metuchen, N.J., pp.189-200

BELOFF, J. (1973b) Psychological Sciences. A Review of Modern Psychology. Crosby, Lockwood, Staples, London.

BELOFF, J. (1974a) (Ed.) New Directions in Parapsychology. Elek Science, London.

BELOFF, J. (1974b) Letter in "The Geller Correspondence", New Scientist, Vol. 64, no.921, 31st October, p.357

BELOFF, J. (1980) Seven evidential experiments (with commentaries from invited commentators and a reply by the author). Zetetic Scholar, Number 6, pp.91-120

BELOFF, J. (1981) The fingerprints of psi. Paper presented at the Fifth International Conference of the Society for Psychical Research, Bristol, England, 10th-12th April.

BELOFF, J. & BATE, D. (1970) Research report for the year 1968-69. University of Edinburgh Parapsychology Unit. Journal of the Society for Psychical Research, Vol.45, no.744, pp.297-301

BELOFF, J. & MANDLEBERG, I. (1967) An attempted validation of the 'Waiting Technique'. Journal of the Society for Psychical Research, Vol.44, no.732, pp.82-88

BENASSI, V.A., REYNOLDS, C. and SINGER, B. (1981) Occult beliefs: seeing is believing. Journal for the Scientific Study of Religion

BESTERMAN, T. (1933) An experiment in 'clairvoyance' with M. Stefan Ossowiecki. Proceedings of the Society for Psychical Research, Vol.41, part 132, pp.345-351

BEVAN, J.M. (1947a) ESP tests in light and darkness. Journal of Parapsychology, Vol.11, no.2, pp.76-89

BEVAN, J.M. (1947b) The relation of attitude to success in ESP scoring. Journal of Parapsychology, Vol.11, pp.296-309

BHADRA, B.H. (1966) The relationship of test scores to belief in ESP. Journal of Parapsychology, Vol.30, no.1, pp.1-17

BLACKMORE, S. (1980) Exrasensory perception as a cognitive process. Unpublished Ph.D. thesis, University of Surrey, England.

BORZYMOWSKI, A. (1962) Parapsychology in Poland: a historical survey. International Journal of Parapsychology, Vol.4, no.4, pp.59-74

BORZYMOWSKI, A. (1965) Experiments with Ossowiecki. International Journal of Parapsychology, Vol.7, no.3, pp.259-284

BRAUD, W., DAVIS, G. and WOOD, R. (1979) Experiments with Matthew Manning. Journal of the Society for Psychical Research, Vol.50, no.782, pp.199-223

BRINKMAN, W. & van HILTEN, W. (1972) Een experimenteel onderzoek naar de effecten van varierende stimulus-inhoud op ESP. (An experimental investigation on the effects on ESP of stimulus contents variance.) SCEPP Bulletin, Vol.2, pp.25-34. Abstract in Journal of Parapsychology, 1972, Vol.36, no.2, p.179

BRUCK, C. (1925) Experimentelle Telepathie. Neu Versuche zur Telepathischen Uebertragung von Zeichnungen. Julius Puttmann Verlag, Stuttgart.

BURDICK, D.S. & KELLY, E.F. (1977) Statistical methods in parapsychological research. In B.B. Wolman (Ed.), Handbook of Parapsychology. Van Nostrand Reinhold, New York, pp.81-130

CADORET, R.J. (1953) The effect of amytal and dexadrine on ESP performance. Journal of Parapsychology, Vol.17, no.4, pp.259-274

CARINGTON, W.W. (1940a) Experiments on the paranormal cognition of drawings, I. Proceedings of the Society for Psychical Research, Vol.46, part 162, pp.34-151. Also in Journal of Parapsychology, 1940, Vol.4, no.1, pp.1-129

CARINGTON, W.W. (1940b) Some observations on the experiments with drawings. Journal of Parapsychology, Vol.4, no.1, pp.130-134

CARINGTON, W.W. (1940c) Reply to Mr. Stevens's criticisms. Proceedings of the Society for Psychical Research, Vol.46, part 163, pp.261-264

CARINGTON, W.W. (1941) Experiments on the paranormal cognition of drawings. Proceedings of the Society for Psychical Research, Vol.46, part 164, pp.277-344

CARINGTON, W.W. (1944a) Experiments on the paranormal cognition of drawings. III. Steps in the development of a repeatable technique. Proceedings of the American Society for Psychical Research, Vol.24, pp.3-107

CARINGTON, W.W. (1944b) Experiments on the paranormal cognition of drawings, IV. Proceedings of the Society for Psychical Research, Vol.47, part 168, pp.155-228

CARINGTON, W.W. (1945) Telepathy. An Outline of its Facts, Theory and Implications. Methuen, London.

CARINGTON, W.W. & HEYWOOD, R. (1944) Some positive results from a group of small experiments. Proceeedings of the Society for Psychical Research, Vol.47, part 168, pp.229-236

CARPENTER, J.C. (1965) An exploratory test of ESP in relation to anxiety proneness. In J.B. Rhine (Ed.), Parapsychology: From Duke to FRNM. Parapsychology Press, Durham, N.C.

CARPENTER, J.C. (1977) Unpublished study, mentioned in "Intra-subject and subject-agent effects in ESP experiments" (p.255). In B.B. Wolman (Ed.), Handbook of Parapsychology, Van Nostrand Reinhold, New York.

CATTELL, R.B., EBER, H.W. and TATSUOKA, M.M. (1970) Handbook for the Sixteen Personality Factor Questionnaire (16PF). Institute for Personality and Ability Testing, Champaign, Ill.

CHAPMAN, D.W. (1934) The statistics of the Method of Correct Matchings. American Journal of Psychology, Vol.46, pp.287-298

CHILD, I.L. & LEVI, A. (1979) Psi-missing in free-response settings. Journal of the American Society for Psychical Research, Vol.73, no.3, pp.273-289

COOK, D.A. (1980) Vision in critical balance: the teachings of Gardner Murphy. Journal of the American Society for Psychical Research, Vol.74, no.1, pp.94-101

DELANOY, D.L. (1981) The training of ESP in the ganzfeld. Paper presented at the Fifth International Society for Psychical Research Conference, Bristol, England, 10th-12th April.

DELIN, P.S. (1977) Two modes of ESP: A re-evaluation of the Stanford Research Institute experiment with Geller. Unpublished manuscript. Abstract in Journal of Parapsychology, Vol.41, no.1, pp.77-78

DESSOIR, M. (1886) Experiments in muscle-reading and thought-transference. Proceedings of the Society for Psychical Research, Vol.4, part 10, pp.111-126

DESSOIR, M. (1888) Experiments in thought-transference. Proceedings of the Society for Psychical Research, Vol.5, part 13, pp.355-357

DINGWALL, E.J. (1924) An experiment with the Polish medium Stephan Ossowiecki. Journal of the Society for Psychical Research, Vol.21, no.405, pp.259-263

DRAKE, L.E. (1946) A Social I.E. Scale for the MMPI. Journal of Applied Psychology, Vol.30, pp.51-54

EHRENWALD, J. (1978) Einstein skeptical of ESP? Postscript to a correspondence. Journal of Parapsychology, Vol.42, no.2, pp.137-142

EHRENWALD, J. (1980) In "Tributes honoring the memory of Gardner Murphy." Journal of the American Society for Psychical Research, Vol.74, no.1, p.116

EISENBERG, H. & DONDERI, D.C. (1979) Telepathic transfer of emotional information in humans. Journal of Psychology, Vol.103, no.1, pp.19-43

EISENBUD, J. (1965) Perception of subliminal visual stimuli in relation to ESP. International Journal of Parapsychology, Vol.7, no.2, pp.161-181

ELKISCH, P. (1945) Children's drawings in a projective technique. Psychological Monographs, Vol.58, pp.1-31

ELLSON, D.G. (1940) A criticism of Dr. Pratt's use of Chapman's "Statistics of the Method of Correct Matchings". Journal of Parapsychology, Vol.4, no.2, pp.329-336

EVANS, C. (1973) Parapsychology -- what the questionnaire revealed. New Scientist, Vol.57, no.830, 25th January, p.209

EYSENCK, H.J. (1967) Personality and extra-sensory perception. Journal of the Society for Psychical Research, Vol.44, no.732, pp.55-71

EYSENCK, H.J. & EYSENCK, S.B.G. (1964) Manual of the Eysenck Personality Inventory. University of London Press, London.

FESTINGER, L. (1957) A Theory of Cognitive Dissonance. Stanford University Press, Stanford, California.

FISK, G.W. (1960) The Rhodes Experiment. Linkage in Extrasensory Perception, by M.C. Marsh. Journal of the Society for Psychical Research, Vol.40, no.703, pp.219-239

GIBSON, H.B. (1979) The 'royal nonesuch' of parapsychology. Bulletin of the British Psychological Society, Vol.32, February, pp.65-67

GREENWOOD, J.A. (1943) A preferential matching problem. Psychometrika, Vol.8, no.3, pp.185-191

GREVILLE, T.N.E. (1944) On multiple matching with one variable deck. Annals of Mathematical Statistics, Vol.15, pp.432-434

GURNEY, E., MYERS, F.W.H. and BARRETT, W.F. (1883) Second report on thought-transference. Proceedings of the Society for Psychical Research, Vol.1, part 2, pp.70-98

GURNEY, E., Myers, F.W.H. and BARRETT, W.F. (1884a) Fourth report of the committee on thought-transference. Proceedings of the Society for Psychical Research, Vol.2, part 5 , pp.1-11

GURNEY, E., MYERS, F.W.H. and BARRETT, W.F. (1884b) Diagrams illustrative of thought-transference. Proceedings of the Society for Psychical Research, Vol.2, part 6, pp.207-216

GURNEY, E., MYERS, F.W.H. and PODMORE, F. (1886) Phantasms of the Living. London.

GURNEY, E., MYERS, F.W.H., PODMORE, F. and BARRETT, W.F. (1883) Third report on thought-transference. Proceedings of the Society for Psychical Research, Vol.1, part 3, pp.161-216

GUTHRIE, M. (1884) An account of some experiments in thought-transference. Proceedings of the Society for Psychical Research, Vol.2, part 5, pp.24-42

GUTHRIE, M. (1885) Further report on experiments in thought-transference at Liverpool. Proceedings of the Society for Psychical Research, Vol.3, part 9, pp.424-452

HANKS, P. (1971) (Ed.) Encyclopedic World Dictionary. Paul Hamlyn, London.

HANLON, J. (1974) Uri Geller and science. New Scientist, Vol.64, no.919, 17th October, pp.170-185

HARALDSSON, E. & JOHNSON, M.U. (1979) ESP and the defense mechanism test (DMT), Icelandic study no. III. A case of experimenter effect? European Journal of Parapsychology, Vol.3, no.1, pp.11-20

HARDING, S. (1980) Meditation, clairvoyant ability and psychological adjustment. Unpublished M.A. thesis, Department of Psychology, University of Edinburgh.

HARDING, S.E. & THALBOURNE, M.A. (1981) Transcendental Meditation, clairvoyant ability and psychological adjustment. In W.G. Roll & J. Beloff (Eds.), Research in Parapsychology 1980. Scarecrow Press, Metuchen, N.J. & London, in press.

HARDY, A., HARVIE, R. and KOESTLER, A. (1973) The Challenge of Chance. Hutchison, London.

HASTED, J. (1981) The Metal-benders. Routledge & Kegan Paul, London.

HAYS, W.L. (1963) Statistics for Psychologists. Holt, Rinehart and Winston, New York.

HETTINGER, J. (1941) Exploring the UltraPerceptive Faculty. Rider & Co., London.

HITCHING, F. (1980) Psi in the laboratory: 12 crucial findings. SPR Study Guide No.1. Society for Psychical Research, London.

HOFF, R. Van't et al. (1972) De invloed van een viertal variabelen op de telepathische waarneming, gemeten met behulp van kloktijden (The influence of four variables on telepathic perception, tested by the use of clock-face targets.) SCEPP Bulletin, Vol.2, pp.54-66. Abstract in Journal of Parapsychology, 1972, Vol.36, no.3, p.252

HONORTON, C. (1977) Psi and internal attention states. In B.B. Wolman (Ed.), Handbook of Parapsychology. Van Nostrand Reinhold, New York, pp.435-472

HUMPHREY, B.M. (1946a) Success in ESP as related to form of response drawings: I. Clairvoyance experiments. Journal of Parapsychology, Vol.10, no.2, pp.78-106

HUMPHREY, B.M. (1946b) Success in ESP as related to form of response drawings: II. GESP experiments. Journal of Parapsychology, Vol.10, no.3, pp.181-196

JONES, J.N. & FEATHER, S.R. (1969) Relationship between reports of psi experiences and subject variance. Journal of Parapsychology, Vol.33, pp.311-319

KAHN, S.D. (1952) Studies in extrasensory perception: Experiments utilizing an electronic scoring device. Proceedings of the American Society for Psychical Research, Vol.25, pp.1-48

KANTHAMANI, B.K. & RAO, K.R. (1973) Personality characteristics of ESP subjects: V. Graphic expansiveness and ESP. Journal of Parapsychology, Vol.37, no.2, pp.119-129

KENDALL, M.G. & BABINGTON SMITH, B. (1939) Tables of random sampling numbers. Tracts for Computers, Vol. 24, London University, University College Department of Statistics.

KENNEDY, J.E. (1979) Methodological problems in free-response ESP experiments. Journal of the American Society for Psychical Research, Vol.73, no.1, pp.1-15

KIM, JAE-ON (1975) Factor analysis. Chapter 24 in N.H. Nie et al., (Eds.), SPSS: Statistical Package for the Social Sciences. 2nd edition, McGraw-Hill, New York.

KIM, JAE-ON & KOHOUT, F.J. (1975) Multiple regression analysis: subprogram REGRESSION. Chapter 20 in N.H. Nie et al. (Eds.), SPSS: Statistical Package for the Social Sciences. 2nd edition, McGraw-Hill, New York.

KNOWLES, E.A.G. (1968) Psi dexterity in a mixing experiment. In J.B. Rhine & R. Brier (Eds.), Parapsychology Today. Citadel, New York, pp.29-34

KRIPPNER, S. (1977) (Ed.) Advances in Parapsychological Research. 1. Psychokinesis. Plenum Press, New York.

KRIPPNER, S. (1978) (Ed.) Advances in Parapsychological Research. 2. Extrasensory Perception. Plenum Press, New York.

KUHN, T.S. (1970) The Structure of Scientific Revolutions. Second edition, enlarged. University of Chicago Press, Chicago.

LEHMANN, A. & HANSEN, F.C.C. (1895) Uber unwillkurliches Flustern (On involuntary whispering.) Philosophische Studien, Vol.11, part 4.

LODGE, O.J. (1884a) An experiment in thought-transference. Letter to Nature, Vol.30, 12th June, p.145

LODGE, O.J. (1884b) An account of some experiments in thought-transference. Proceedings of the Society for Psychical Research, Vol.2, part 6, pp.189-200

LODGE, O.J. (1892) Some recent thought-transference experiments. Proceedings of the Society for Psychical Research, Vol.7, part 20, pp.374-382

MAKRIS, M. (1979) Preliminary report on the findings from a remote viewing investigation. Unpublished Honours thesis, Department of Psychology, University of Adelaide, South Australia

MARSH, M.C. (1958) Linkage in extra-sensory perception. Unpublished Ph.D. dissertation, Rhodes University, Grahamstown, South Africa.

MARSH, M.C. (1962) Three ESP experiments using drawings as target material. Publications of the South African Society for Psychical Research, Vol.5, pp.4-15

McMAHAN, E.A. (1946) An experiment in pure telepathy. Journal of Parapsychology, Vol.10, no.4, pp.224-242

MILLAR, B. (1979) The distribution of psi. European Journal of Parapsychology, Vol.3, no.1, pp.78-110

MITTENECKER, E. & SCHULTER, G. (1978) Bericht uber einen Telepathie-Versuch. Berichte aus dem Institut fur Psychologie, der Universitat Graz, December.

MORRIS, R.L. (1972) An exact method for evaluating preferentially -matched free-response material. Journal of the American Society for Psychical Research, Vol.66, no.4, pp.401-407

MOSS, T. & GENGERELLI, J.A. (1968) ESP effects generated by affective states. Journal of Parapsychology, Vol.32, no.2, pp.90-100

MURPHY, G. (1943) Psychical phenomena and human needs. Journal of the American Society for Psychical Research, Vol.37, no.4, pp.163-191

MUSSO, J.R. (1965) ESP experiments with primary school children. Journal of Parapsychology, Vol.29, no.2, pp.115-121

MUSSO, J.R. & GRANERO, M. (1973) An ESP drawing experiment with a high-scoring subject. Journal of Parapsychology, Vol.37, no.1, pp.13-36

MUSSO, J.R. & GRANERO, M. (1980) Comments on J. Beloff's "Seven evidential experiments", in Zetetic Scholar, Number 6, pp.100-103

MYERS, F.W.H. (1903) Human Personality. (2 volumes) Longmans, Green & Co., New York.

NASH, C.B. & RICHARDS, A. (1947) Comparison of two distances in PK tests. Journal of Parapsychology, Vol.11, pp.269-282

NIE, N.H., HULL, C.H., JENKINS, J.G., STEINBRENNER, K. and BENT, D.H. (1975) SPSS: Statistical Package for the Social Sciences. 2nd edition, McGraw-Hill, New York.

OSSOWIECKI, S. (1933) Swiat Mego Ducha (The World of my Spirit). Warsaw.

O'SULLIVAN, M. & GUILFORD, J.P. (1966) Six Factor Tests of Social Intelligence. Manual of Instructions and Interpretations. Sheridan Psychological Services, Beverly Hills, Ca.

PALMER, J. (1971) Scoring in ESP tests as a function of belief in ESP. Part I. The Sheep-Goat Effect. Journal of the American Society for Psychical Research, Vol.65, no.4, pp.373-408

PALMER, J. (1972) Scoring in ESP tests as a function of belief in ESP. Part II. Beyond the Sheep-Goat Effect. Journal of the American Society for Psychical Research, Vol.66, no.1, pp.1-26

PALMER, J. (1975) Three models of psi test performance. Journal of the American Society for Psychical Research, Vol.69, no.4, pp.333-339

PALMER, J. (1977) Attitudes and personality traits in experimental ESP research. In B.B. Wolman (Ed.), Handbook of Parapsychology, Van Nostrand Reinhold, New York, pp.175-201

PALMER, J. (1978) Extrasensory perception: research findings. In S. Krippner (Ed.), Advances in Parapsychological Research. 2. Extrasensory Perception. Plenum Press, New York, pp.59-243

PALMER, J. (1980) Parapsychology as a probabilistic science: facing the implications. Presidential Address to the Twenty-Second Annual Convention of the Parapsychological Association. In W.G. Roll (Ed.), Research in Parapsychology, 1979. Scarecrow Press, Metuchen, N.J. & London, pp.189-215

PALMER, J. & MILLER, A. (1972) Monetary incentive and the Sheep-Goat effect. In W.G. Roll, R.L. Morris and J.D. Morris (Eds.), Proceedings of the Parapsychological Association Number 7, 1970. Parapsychological Association, Durham, N.C., pp.11-12

PICKERING, W.H. (1885) Thought-transference in Boston. Science, Vol.6, no.126, 3rd July, pp.8-9

PRATT, J.G. (1937) The work of Dr. C. Hilton Rice in extra-sensory perception. Journal of Parapsychology, Vol.1, pp.239-259

PRATT, J.G. (1940) Comment on Dr. Ellson's criticism. Journal of Parapsychology, Vol.4, no.2, p.337

PRATT, J.G. (1966) New ESP tests with Mrs. Gloria Stewart. Journal of the American Society for Psychical Research, Vol.60, no.4, pp.321-339

PRINCE, W.F. (1932) The Sinclair experiments demonstrating telepathy. Bulletin XVI of the Boston Society for Psychical Research, part 1, pp.1-135. Also reprinted as an addendum to U. Sinclair's Mental Radio, Charles C. Thomas, Springfield, Ill., revised second printing, 1962, pp.149-235

PUTHOFF, H.E., TARG, R. and TART, C.T. (1980) Resolution in remote viewing studies: mini-targets. In W.G. Roll (Ed.), Research in Parapsychology 1979. Scarecrow Press, Metuchen, N.J. & London, pp.120-122.

RAO, K.R. (1965) The bidirectionality of psi. Journal of Parapsychology, Vol.29, pp.230-250

RAO, K.R. (1974) Psi and personality. Chapter 3 in J. Beloff (Ed.), New Directions in Parapsychology. Elek Science, London, pp.60-76

RAWSON, H.G. (1895) Experiments in thought-transference. Proceedings of the Society for Psychical Research, Vol.11, part 27, pp.1-17

RHINE, J.B. (1969) Psi-missing re-examined. Journal of Parapsychology, Vol.33, no.1, pp.1-38

RHINE, J.B. (1974) Telepathy and other untestable hypotheses. Journal of Parapsychology, Vol.38, no.2, pp.137-153

RHINE, L.E. (1967) Toward understanding psi-missing. In W.G. Roll (Ed.), Proceedings of the Parapsychological Association,

Number 2, 1965. Parapsychological Association, Durham, N.C., pp.61-78. Also in Journal of Parapsychology, Vol.29, 1965, pp.259-274.

RICE, G.E. & TOWNSEND, J. (1962) Agent-percipient relationship and GESP performance. Journal of Parapsychology, Vol.26, pp.211-217

RICE, G.E., WILLISS, D., LAFFERTY, C., LITTLE, J. and MAULDIN, C.H. (1966) Emotional closeness, communication of affect, and ESP. Journal of Parapsychology, Vol.30, no.4, pp.282-283

RICHET, C. (1888) Relation de diverses experiences sur la transmission mentale, la lucidite, et autres phenomenes non explicables par les donnees scientifiques actuelles. Proceedings of the Society for Psychical Research, Vol.5, part 12, pp.18-168

RUSH, J.H. & JENSEN, A. (1949) A reciprocal distance GESP test with drawings. Journal of Parapsychology, Vol.13, pp.122-134

SCHMEIDLER, G.R. (1943) Predicting good and bad scores in a clairvoyance experiment: a preliminary report. Journal of the American Society for Psychical Research, Vol.37, pp.103-110

SCHMEIDLER, G.R. (1945) Separating the sheep from the goats. Journal of the American Society for Psychical Research, Vol.39, pp.47-49

SCHMEIDLER, G.R. (1946) Progress report on further sheep-goat series. Journal of the American Society for Psychical Research, Vol.40, pp.34-35

SCHMEIDLER, G.R. (1960) ESP in relation to Rorschach test evaluation. Parapsychological Monographs No.2. Parapsychology Foundation, New York.

SCHMEIDLER, G.R. (1961) Evidence for two kinds of telepathy. International Journal of Parapsychology, Vol.3, no.3, pp.5-48

SCHMEIDLER, G.R. (1969) (Ed.) Extrasensory Perception. Atherton, New York.

SCHMEIDLER, G.R. & ALLISON, L.W. (1948) A repetition of Carington's experiments with free drawings. Journal of the American Society for Psychical Research, Vol.42, pp.97-107

SCHMIDT, H. (1969) Precognition of a quantum process. Journal of Parapsychology, Vol.33, no.2, pp.99-108

SCHMIDT, H. (1974) Comparison of PK action on two different random number generators. Journal of Parapsychology, Vol.38, no.1, pp.47-55

SCHMIDT, H. & PANTAS, L. (1972) Psi tests with internally different machines. Journal of Parapsychology, Vol.36, no.3, pp.222-232

SCHMOLL, A. (1887) Experiments in thought-transference. Proceedings of the Society for Psychical Research, Vol.4, part 11, pp.324-337

SCHMOLL, A. & MABIRE, J.E. (1888) Experiments in thought-transference. Proceedings of the Society for Psychical Research, Vol.5, part 12, pp.169-215

SCHRENCK-NOTZING, Baron von (1891) Experimental studies in thought-transference. Proceedings of the Society for Psychical Research, Vol. 7, part 18, pp.3-22

SCHOUTEN, S.A. (1979) Analysis of spontaneous cases as reported in "Phantasms of the Living". European Journal of Parapsychology, Vol.2, no.4, pp.408-455

SCHWARZ, B.E. (1963) Psychodynamic Experiments in Telepathy. New Jersey. Reprinted from Corrective Psychiatry and Journal of Social Therapy, Vol.9, no.4. Reviwed by G.W. Fisk in Journal of the Society for Psychical Research, Vol.42, 1964, no.720, pp.309-316

SCOTT, C. (1972) On the evaluation of verbal material in parapsychology: A discussion of Dr. Pratt's monograph. Journal of the Society for Psychical Research, Vol.46, no.752, p.79-90

SIDGWICK, H. (1896) Involuntary whispering considered in relation to experiments in thought-transference. Proceedings of the Society for Psychical Research, Vol.12, part 31, pp.298-315

SIEGEL, S. (1956) Non-parametric Statistics for the Behavioral Sciences. McGraw-Hill, New York.

SINCLAIR, U. (1930) Mental Radio. Charles C. Thomas, Springfield, Ill. Revised second printing, 1962.

SINGER, B. & BENASSI, V.A. (1981) Occult beliefs. American Scientist, Vol.69, January-February, pp.49-55

SMITH, B.M. & HUMPHREY, B.M. (1946) Some personality characteristics related to ESP performance. Journal of Parapsychology, Vol.10, no.4, pp.269-289

SOLFVIN, G.F., KELLY, E.F. and BURDICK, D.S. (1978) Some new methods of analysis for preferential-ranking data. Journal of the American Society for Psychical Research, Vol.72, no.2, pp.93-109

STANFORD, R.G. (1964) Differential position effects for above-chance scoring sheep and goats. Journal of Parapsychology, Vol.28, pp.155-165

STANFORD, R.G. (1978) Toward reinterpreting psi events. Journal of the American Society for Psychical Research, Vol.72, pp.197-214

STANFORD, R.G. & BRIER, R. (1968) Cancellation effects within the test run. In J.B. Rhine & R. Brier (Eds.), Parapsychology Today. Citadel, New York, pp.54-62.

STEVENS, W.L. (1939) Tests of significance for extra-sensory perception data. Psychological Review, Vol.46, pp.142-150

STEVENS, W.L. (1940) On the interpretation of the data of certain experiments in paranormal cognition. Proceedings of the Society for Psychical Research, Vol.46, part 163, pp.256-260

STUART, C.E. (1942) An ESP test with drawings. Journal of Parapsychology, Vol.6, no.1, pp.20-43

STUART, C.E. (1944) The Carington free-drawing approach to the ESP problem. Journal of Parapsychology, Vol.8, pp.127-138

STUART, C.E. (1945a) A classroom ESP experiment with the free response method. Journal of Parapsychology, Vol.9, no.2, pp.92-105

STUART, C.E. (1945b) An ESP experiment with enclosed drawings. Journal of Parapsychology, Vol.9, no.4, pp.278-295

STUART, C.E. (1946) GESP experiments with the free response method. Journal of Parapsychology, Vol.10, no.1, pp.21-35

STUART, C.E. (1947) A second classroom ESP experiment with the free response method. Journal of Parapsychology, Vol.11, no.1, pp.14-25

STUART, C.E., HUMPHREY, B.M., SMITH, B.M. and McMAHAN, E.A. (1947) Personality measurements and ESP tests with cards and drawings. Journal of Parapsychology, Vol.11, no.2, pp.118-146

TARG, R. & PUTHOFF, H. (1974a) Information transmission under conditions of sensory shielding. Nature, Vol.251, 18th October, pp.602-607

TARG, R. & PUTHOFF, H. (1974b) Geller: experimenters reply. In Letters, New Scientist, Vol.64, no.922, 7th November, p.443

TARG, R. & PUTHOFF, H. (1978) Mind Reach. Granada, London.

TARG, R., PUTHOFF, H.E., HUMPHREY, B.S. and TART, C.T. (1980) Investigations of target acquisition. In W.G. Roll (Ed.), Research in Parapsychology 1979. Scarecrow Press, Metuchen, N.J. & London, pp.122-124

TAVES, E. (1945) The construction of an American Catalogue. Journal of the American Society for Psychical Research, Vol. 39, pp.151-156

TAVES, E., MURPHY, G. and DALE, L.A. (1945) American experiments on the paranormal cognition of drawings. Journal of the American Society for Psychical Research, Vol.39, pp.144-150

TERESI, D. (1981) People. Omni, Vol.3, no.7, April, p.118

THALBOURNE, M.A. (1976) Closeness of relationship, and telepathy, personality and social intelligence. Unpublished Honours thesis, Department of Psychology, University of Adelaide, South Australia.

THALBOURNE, M.A. (1979a) A long-distance ESP drawing experiment between Austria and Iceland. Paper presented at the Third Annual Conference of the Society for Psychical Research, Edinburgh, Spring 1979. (Abstract published by the Society for Psychical Research, London, p.22)

THALBOURNE, M.A. (1979b) A more powerful method of evaluating data from free-response experiments. Journal of the Society for Psychical Research, Vol.50, no.780, pp.84-107

THALBOURNE, M.A. (1980) Two long-distance ESP drawing experiments between Austria and Iceland. In W.G. Roll (Ed.), Research in Parapsychology 1979. Scarecrow Press, Metuchen, N.J. & London, pp.174-177.

THALBOURNE, M.A. (1981) Extraversion and the sheep-goat variable: a conceptual replication. Journal of the American Society for Psychical Research, Vol.75, no.2, pp.105-119

THALBOURNE, M.A., BELOFF, J. and DELANOY, D.L. (1982) A test of the "extraverted sheep versus introverted goats" hypothesis. In W.G. Roll & R.L. Morris (Eds.), Research in Parapsychology 1981. Scarecrow Press, Metuchen, N.J. & London.

THALBOURNE, M.A. & HARALDSSON, E. (1980) Personality characteristics of sheep and goats. Personality and Individual Differences, Vol.1, no.2, pp.180-185. Also in W.G. Roll (Ed.), Research in Parapsychology 1979. Scarecrow Press, Metuchen, N.J. & London, pp.100-104

THORPE, L.P., CLARK, W.W. and TIEGS, E.W. (1953) California Test of Personality: Manual. 1953 Revision. All levels, forms AA and BB. California Test Bureau, Los Angeles.

THOULESS, R.H. (1963) Experimental Psychical Research. Penguin, London.

THOULESS, R.H. (1972) From Anecdote to Experiment in Psychical Research. Routledge & Kegan Paul, London.

THOULESS, R.H. & WIESNER, B.P. (1947) The psi processes in normal and 'paranormal' psychology. Proceedings of the Society for Psychical Research, Vol.48, part 174, pp.177-196

TISCHNER, R. (1925) Telepathy and Clairvoyance. Harcourt, Brace & Co., New York.

ULLMAN, M. & KRIPPNER, S. (1970) Dream Studies and Telepathy: An Experimental Approach. Parapsychological Monographs, no.12. Parapsychology Foundation, New York.

USHER, F.L. & BURT, F.P. (1909) Thought-transference. Annals of Psychical Science, Vol.8, pp.561-600

VELLISSARIS, C.N. & VELLISSARIS, C.R. (1977) Similar experience-memory factors and psi scoring. New England Journal of Parapsychology, Vol.1, no.1, pp.4-18

WARCOLLIER, R. (1921) La Telepathie. Alcan, Paris.

WARCOLLIER, R. (1938) Experimental Telepathy. Boston Society for Psychical Research, Boston, Mass.

WARCOLLIER, R. (1948) Mind to Mind. Creative Age Press, New York. New, enlarged edition published by Collier, New York, 1963.

WEST, D.J. (1947) Mass experiments in the psi cognition of drawings. Journal of the Society for Psychical Research, Vol.34, no.635, pp.43-54

WEST, D.J. (1950) ESP performance and the expansion-compression rating. Journal of the Society for Psychical Research, Vol.35, no.660, pp.295-308

WEST, D.J. (1962) Psychical Research Today.

WHEELER, J.A. (1979) Drive the pseudos out of the workshop of science. In M. Gardner, "Quantum theory and quack theory". New York Review of Books, 17th May 1979, pp.40-41

WOLMAN, B.B. (1977) (Ed.) Handbook of Parapsychology. Van Nostrand Reinhold, New York, N.Y.

A G L O S S A R Y O F T E R M S
U S E D I N
P A R A P S Y C H O L O G Y

Compiled by

Michael A. Thalbourne

PREFACE

The meaning of a term is often the key to the proper understanding of some technique or specialized study in the sciences and arts. The science of parapsychology is no exception: over the past century or so it has been continually evolving its own unique vocabulary, and without proper comprehension of these terms the student may have as much difficulty in reading the literature of the subject as if it were written in a foreign language.

Yet it is probably true to say that hitherto there has been no exhaustive glossary of parapsychological terms which was at once inexpensive and readily accessible. To my knowledge, the most comprehensive glossary to date has been the excellent one compiled by Laura Dale and Rhea White, to be found in Benjamin Wolman's Handbook of Parapsychology. But though this handbook is an essential text and reference work for all professional parapsychologists, it is, unfortunately, very expensive (particularly outside the U.S.A.), and its sheer size of nearly one thousand pages makes it rather too hefty a tome to be considered a handy sort of handbook!

But a more cogent reason for bringing out this particular glossary is the fact that, in any evolving science, new terms are continually being added and old ones becoming obsolete, as knowledge grows and concepts change. New phenomena, effects and techniques are described, or older ones are more precisely delineated, and distinctions drawn. One has only to catalogue some of the neologisms that the 1970s witnessed to appreciate this point: terms like "ganzfeld", "remote viewing", "dermo-optical perception", "conformance behaviour", "micro-PK", "psychotronics" -- the list is not quite endless but nonetheless long. In fact, this present glossary contains approximately 80 terms which will not be found in any previous compilation, and perhaps half of them are of very recent vintage. The work of the glossarist will cease only when psychical research has ceased to progress. One can envisage, then, that in the not-too-distant future, it will be appropriate and worthwhile to produce a second, revised edition of this glossary so as to incorporate the latest developments. To that end, I would be most grateful to hear from any reader who has suggestions for improving the present edition, (especially regarding any errors, of which there are undoubtedly some,

though hopefully few), and for any terms -- new or old -- which they think could be usefully included.

Two unique features of the present work should perhaps be noted here, namely, the inclusion of a certain amount of etymological information, as well as references to primary source material. As regards the first feature, I have always, myself, found it useful and interesting to know the derivation of a word: it not only seems to make the term and its meaning more readily remembered but also puts us in mind of the conceptual framework from which it sprang: for example, the knowledge that "hypnagogic" comes from two Greek words, "hypnos" (= "sleep") and "agōgos" (= "leading"), should help demystify the term for non-classicists, even before a more extended definition is given it. Perhaps this hope is an over-optimistic one, born of my penchant for philology, but I trust that not too many readers will see it as an unnecessary encumbrance.

But almost universal consensus can be anticipated for conceding the utility of providing the reader with references to the original source material where the term or expression was first used. Particularly in the case of theoretical models, such as Rex Stanford's 'conformance behaviour model' or Thouless and Wiesner's 'Shin theory', it must surely be invaluable to have information as to where to go to hear the "words of the master", so to speak. I make no claim for completeness in this respect, however, and one can only promise a more detailed treatment if and when feedback from this edition suggests its usefulness.

It remains to discharge a sincere debt of thanks to a number of people: to Kathy Wilson, Deborah Delanoy, Sue Fisher and Dr. Adrian Parker, for reading the manuscript and making helpful suggestions; and most especially to my friend, colleague and supervisor, Dr. John Beloff, whose valuable comments and guidance at every stage of the production of this glossary, have made it a much more erudite work than it would otherwise have been: for his encouragement and his knowledge, he has my enduring gratitude.

M.A.T.
Edinburgh,
February, 1980

SOURCE MATERIAL

ASHBY, R.H. Glossary of Terms, in The Guidebook for the Study of Psychical Research, Rider: London, 1972, pp. 144-157

BELOFF, J. Entries passim, in H.J. Eysenck, W. Arnold, & R. Meili (Eds.) Encyclopædia of Psychology, Search Press: London, 1972

BELOFF, J. Glossary of Statistical Terms and Abbreviations used in the Text, in New Directions in Parapsychology, Elek Science: London, 1974, pp. xiv-xxvi.

DALE, L.A. & WHITE, R.A. Glossary of Terms Found in the Literature of Psychical Research and Parapsychology, in B.B. Wolman (Ed.) Handbook of Parapsychology, Van Nostrand Reinhold: New York, N.Y., 1977, pp. 921-936

HANKS, P. (Ed.) Encyclopedic World Dictionary, Hamlyn: London, 1971

JOURNAL OF PARAPSYCHOLOGY Glossary, published in every issue, Parapsychology Press: Durham, N.C.

LIDDELL & SCOTT Greek-English Lexicon, Clarendon Press: Oxford, 1953

NASH, C.B. Appendix II, in Science of Psi, Charles C. Thomas: Springfield, Ill., 1978, pp. 237-249

RAO, K.R. Glossary, in Experimental Parapsychology, Charles C. Thomas: Springfield, Ill., 1966, pp.244-246

RHINE, J.B. & BRIER, R. (Eds.) Glossary, in Parapsychology Today, Citadel: New York, N.Y., 1968, pp. 267-273

RHINE, J.B. & PRATT, J.G. Glossary, in Parapsychology. Frontier Science of the Mind, Charles C. Thomas: Springfield, Ill., 1957, pp. 207-210

WARCOLLIER, R. Glossary, in Mind to Mind, Collier: New York, N.Y., 1963, pp. 106-109

WEST, D.J. Appendix V. Glossary of Terms used in ESP Research, in Tests for Extrasensory Perception, Revised edition, Society for Psychical Research: London, 1954, pp. 25-27

TABLE OF ABBREVIATIONS

A. In whole or in part, quoted verbatim from Ashby (1972).

B,a. In whole or in part, quoted verbatim from Beloff (1972).

B,b. In whole or in part, quoted verbatim from Beloff (1974).

Cf. (L. confer) compare.

D & W. In whole or in part, quoted verbatim from Dale & White (1977).

der. derived from.

EJP. European Journal of Parapsychology.

F. French.

fm. from.

Gk. Greek.

JASPR. Journal of the American Society for Psychical Research.

JP. Journal of Parapsychology.

JP. In whole or in part, quoted verbatim from the Journal of Parapsychology.

JSPR. Journal of the Society for Psychical Research.

L. Latin.

N. In whole or in part, quoted verbatim from Nash (1978).

PSPR. Proceedings of the Society for Psychical Research.

W. In whole or in part, quoted verbatim from Warcollier (1963)

NOTES FOR THE USER

(i) Word order. Entries are listed in simple alphabetic order; this is true also for acronyms, such as "MOBIA" and "NDE": thus, for example, "RNG" is to be found between "REVERSAL EFFECT" and "RSPK": there is no separate subsection of acronyms preceding the first "full" word (viz. "RADIAESTHESIA"), in contrast to such previous glossaries as that by Dale & White (1977). Terms which begin with, or consist of, Greek letters are to be found in the same position that they would occupy if they were transliterated, but preceding their Roman counterparts: thus, for example, "ΨΥ" precedes "PSI-

GAMMA", but follows "PSI-FIELD HYPOTHESIS."

(ii) Use of italics. If a word is italicized, it is either (i) a non-English phrase, (ii) a book title, (iii) a word simply being given emphasis, or (iv) a term which has its own definition elsewhere in the glossary, and which the reader should consult if in doubt as to its technical meaning.

(iii) Primary references. Occasionally, instead of a full reference to a primary source being given, the author's name will be given simply with a publication year in brackets, an example being "Humphrey (1946)": this is merely a space-saving device, and implies that the full reference has been given elsewhere in the glossary; the reader should be able to locate the full reference by consulting the definitions for the italicized (i.e. cross-referenced) terms or those terms which the reader is instructed to "see also" or "compare".

(iv) Acknowledgements to other source glossaries. The reader will often notice an abbreviation such as "(D & W)" at the end of the definition, and this indicates that all or part of that definition has been quoted virtually verbatim ; the Table of Abbreviations (above) lists the authors in question, and the full reference may then be found in the list of Source Material (also above).

(v) Use of the pronoun "they". An idiosyncracy of the present author's grammatical style is the avoidance of such clumsy and male-oriented expressions as "he or she", in favour of the quasi-singular use of the pronoun "they" and its inflections, as for example in the sentence "Whenever a person is kind enough to participate in an experiment of mine, I always try to give them feedback regarding their results".

ABSENT HEALING. See under HEALING, PSYCHIC.

ABSENT SITTING. See PROXY SITTING.

ACTIVE-AGENT TELEPATHY. Term referring to instances of telepathy in which the agent, rather than being simply the person whose mental states are passively scanned, seems to play a more active role in influencing mental or behavioural changes in the target person. See also (K)-TELEPATHY; MENTAL OR BEHAVIOURAL INFLUENCE OF AN AGENT; TELERGY.

AFFECTABILITY. Concept introduced by Charles Stuart ("An analysis to determine a test predictive of extra-chance scoring in card-calling tests", J.P., 1941, 5, 99-137) originally meaning "susceptibility to feedback" in the context of an ESP card-guessing test: "affectable" subjects consistently alter their estimate of what their score will be on the next run, in line with their score on the previous run; "unaffectable" subjects are those whose estimates are not affected by such feedback. Later ("An Interest Inventory relation to ESP scores", J.P., 1946, 10, 154-161), Stuart applied the term "affectable" to subjects who are relatively extreme in the degree to which they express like or dislike towards a number of possible interests, whereas "unaffectable" subjects are relatively indifferent to a large number of interests; affectability can supposedly be be measured by means of the Stuart Interest Inventory. On tests of ESP, unaffectable subjects have been observed to score higher than those who are affectable.

AGENT. In a test of GESP, the individual (human or animal) who looks at the information constituting the target and who is said to "send" or "transmit" that information to a percipient; in a test of telepathy, and in cases of spontaneous ESP, the individual about whose mental states information is acquired by a percipient; very occasionally, the term refers to the subject in a test of psychokinesis. (fm. L. agens, der. agere, "to drive, do")

ALLOBIOFEEDBACK. See under BIOFEEDBACK.

ALPHA. (i) In the context of statistical analysis: see under SIGNIFICANCE; (ii) In the context of brain science: a distinctive brain-rhythm or brain-wave which occurs mainly in the occipital region of the cortex, and which is correlated, on the psychological level, with feelings of drowsiness, relaxation and disengaged attention on the part of the subject; it is of relatively high amplitude, and

has a frequency range of between 8 and 12 Hz (cycles per second); of parapsychological interest as a possible physiological indicator of a psi-conducive condition in the subject. (Gk. alpha, first letter of the Greek alphabet, symbolized ∝)

ALTERED STATE(S) OF CONSCIOUSNESS (ASC). Expression coined by Charles Tart which can refer to virtually any mental state differing from that of the normal waking condition; of parapsychological interest as possibly psi-conducive states; they include dreaming, hypnosis, trance, meditation of the yoga or Zen tradition, the hypnagogic state induced by the ganzfeld, and drug-induced states.

ANPSI. Psi ability in non-human animals. (Contraction of "animal psi")

APPARITION. A visual experience in which there appears to be present a person or animal (deceased or living) who is in fact out of sensory range of the experient; often associated with spontaneous ESP, e.g. in connection with an agent who is dying or undergoing some other crisis (in which case it is termed a "crisis apparition"), or in connection with a haunting (in which case it is likely to be called a "ghost").

APPORT. A physical object alleged to have been paranormally transported into a closed space or container, suggesting the passage of "matter through matter", i.e. through intervening material objects. (fm. L. apportare, "to carry to (a place)")

AROUND-THE-DIE TECHNIQUE. A technique of testing for psychokinesis in which each face of a die is designated as the target an equal number of times, this in order to counteract the influence of any bias present in the dice used.

ARRIVAL CASE. A type of spontaneously occurring ESP in which the percipient has a strong but inexplicable feeling that they are going to meet a specific person, whereupon shortly afterwards that person does indeed arrive.

ASC. See ALTERED STATE(S) OF CONSCIOUSNESS.

ASSOCIATION THEORY OF TELEPATHY. A theory of telepathy proposed by Whately Carington (Telepathy. An outline of its facts, theory, and implications, Methuen, London, 1945) based on an extension of the psychological Law of Association: it postulates that if two ideas, A and B, are presented together, or in close succession, to any mind M, and if subsequently one of them is re-presented to that mind, or presented to any other mind M', then the other idea is more likely to accompany or closely follow it, in the mind to which it is so presented or

re-presented, than if the two ideas had not been presented together to the mind M. See also K-IDEA; K-OBJECT.

ASTRAL BODY. Sometimes also termed "etheric body"; said to be an exact, quasi-physical replica or "double" of the individual physical body, which can separate itself from the physical body, either temporarily, as in dreaming or out-of-the-body experience, or permanently at the moment of death. (fm. L. astralis, der. astrum, "star", der. Gk. astron)

ASTRAL PROJECTION. Term coined by Oliver Fox to describe the temporary separation of the so-called "astral body" from the physical body. See also OUT-OF-THE-BODY EXPERIENCE; TRAVELLING CLAIRVOYANCE

ASTRAL TRAVEL. See ASTRAL PROJECTION.

AUGURY. See DIVINATION. (fm. L. augurium)

AURA. A putative field of subtle, multi-coloured, luminous radiations surrounding living bodies as a halo or cocoon; the term is occasionally used to refer to the normal electromagnetic field-forces surrounding the body. See also AURA READING. (L., fm. Gk., "breath of air")

AURA READING. A description of a target person (their physical or mental state, personality, problems, etc.), supposedly inferred from the perceived characteristics (colour, form, density, etc.) of their aura by a sensitive. Hence "aura reader", a person who gives such a reading.

AUTHENTICATION. The verification of the facts and details surrounding the occurrence of a spontaneous psi phenomenon by independent statements from witnesses, written accounts, or other corroboratory material supporting the report made by the percipient concerning their experience and the events to which it was apparently related. See also CORROBORATOR.

AUTOLEVITATION. See under LEVITATION.

AUTOMATIC WRITING. A phenomenon in which a person's hand writes spontaneously, producing meaningful sentences, but without that person being consciously aware of what is being written. See also AUTOMATISM. (B,a)

AUTOMATISM. Any complex sensory or motor activity carried out by a person without their conscious awareness or volition, and usually while in a dissociated

state; examples are sleep-walking, automatic writing. Hence, "automatist", a person who experiences or practises automatism.

AUTOMATIST. See under AUTOMATISM.

AUTOSCOPY. An act or hallucination in which a person sees their "double", or their own physical body, as if from a point outside the body. See also ASTRAL BODY; OUT-OF-THE-BODY EXPERIENCE. (fm. Gk. autos, "self", + skopia, "a watching", der. skopein, "to look at")

BACKWARD CAUSATION. See RETROACTIVE CAUSATION.

BACKWARD DISPLACEMENT. See under DISPLACEMENT.

BASIC TECHNIQUE (BT). A technique for testing clairvoyance in which the experimenter lays aside each target card as it is guessed by the subject. The correctness of each call is not ascertained until the end of the run.

BIDIRECTIONALITY OF PSI. Term introduced by Ramakrishna Rao ("The bidirectionality of psi", J.P., 1965, 29, 230-250) to refer to the fact that psi may manifest itself in either a positive (above chance) direction or a negative (below chance) direction. See also PSI-HITTING; PSI-MISSING; DIFFERENTIAL EFFECT.

BILLET READING. A test in which the sitter writes a question on a piece of paper and seals it in an envelope; the medium then answers the question, and sometimes gives additional information relevant to the sitter, purportedly by paranormal means. Cf. TOKEN-OBJECT READING. (fm. F., blend of bille, "a writing", and bullette, "certificate") (D & W)

BILOCATION. The alleged phenomenon in which a person is seen in two different geographical locations at the same time.

BIOCOMMUNICATION. Term used for "telepathy" in the literature of Soviet and Eastern European parapsychology. (fm. Gk. bios, "life", + communication) Cf. BIOINFORMATION.

BIOFEEDBACK. A technique which enables a person to monitor ongoing changes in one of their own physiological processes; as a result of such information, the individual can acquire some degree of control in regulating internal processes normally outside the range of voluntary influence; of parapsychological interest

mainly in connection with altered states of consciousness and with the possibility of controlling the incidence of alpha. Hence, "allobiofeedback", a term used by William Braud ("Allobiofeedback: Immediate feedback for a psychokinetic influence upon another person's physiology", Paper presented at the Twentieth Annual Convention of the Parapsychological Association, Washington D.C., August 1977) to denote the situation where one subject, A, is attempting to influence, psychokinetically, the physiological processes of another subject,B, aided by biofeedback to A concerning those processes in B. (fm. Gk. allos, "other", + biofeedback)

BIOINFORMATION. Term used for "extrasensory perception" in the literature of Soviet and Eastern European parapsychology. Cf. BIOCOMMUNICATION.

BIOPLASMA. Term used in Soviet parapsychology to refer to a hypothetical counterpart of the living body, composed of a mixture of electrons, photons and other forms of energy, and involved in extrasensory perception, psychokinesis and Kirlian photography. (fm. Gk. bios, "life", + plasma, "something formed or moulded")

BLIND. (Especially of a judge or other participant in an experiment), without knowledge of cues or information which would reveal the true target or its relationship to the responses in a test of psi. Hence also "double blind", term applied to a situation where not only the judge or subject but also the experimenter or other test-administrator is kept in ignorance of cues or information which would reveal the true target or its relationship to the responses in a test of psi.

BLIND-MATCHING (BM) TECHNIQUE. A method of testing for clairvoyance in which the subject, holding the pack of ESP cards face down, sorts them into five piles in an attempt to match concealed key cards representing the five different symbols.

BM. See BLIND-MATCHING (BM) TECHNIQUE.

BT. See BASIC TECHNIQUE.

BOOK TEST. A test for survival sometimes conducted during a sitting in which an attempt is made to exclude telepathy between medium and sitter by having the communicator transmit a message referring to topics on specific pages of a book that the medium could not have seen normally. See also NEWSPAPER TEST. (D & W)

CABINET. An enclosed space, generally surrounded by curtains, which most physical mediums claim is necessary in order that they can condense the "psychic force" by means of which they produce physical phenomena.

CALL. (As a noun) the overt response made by the subject in guessing the target in a test of ESP; (as a verb), to make a response or call.

CANCELLATION EFFECT. A phenomenon wherein unusually high scores produced in one segment of an experimental unit (such as a run, or a series of runs), when added to exceptionally low scores produced in another segment of that unit, are "cancelled out", in that the overall average success rate for the whole unit is near mean chance expectation.

CHAIR TEST. A test for precognition, associated especially with the Dutch sensitive Gerard Croiset but first demonstrated by Pascal Forthuny, a French psychic, in which a chair is randomly selected from all those set up for a later public meeting, and the percipient describes the appearance, characteristics and events in the life of a person, unknown to them, who will later attend that meeting and occupy that chair.

CHANCE. The complex of undefined causal factors irrelevant to the purpose at hand, often spoken of as if it were a single, real agency. Sometimes, the term is a short-hand expression for "mean chance expectation", as in "deviation from chance". (JP)

CHANGE EFFECT. Term used by Robert Thouless to indicate that a change of conditions in an experimental task is often associated with a temporary drop in scoring level. (D & W)

CHECKER. The person whose role it is to compare the targets with the responses made to them in a test of psi, in order to calculate the subject's score.

CHI-SQUARE (χ^2). A widely used test of statistical significance, applicable when the data consist of frequency counts (as when the data are divided into discrete categories); it is a number greater than or equal to zero, representing a sum of arithmetic quantities, each of which is the square of the difference between observed and expected frequencies in each category, this difference being divided by the expected frequency: $\chi^2 = \Sigma\,(o - e)^2/e$; it can also refer to a sum of the squares of a number of critical ratios. (Gk. chi, twenty-second

letter of the Greek alphabet, symbolized χ)

CIPHER TEST FOR SURVIVAL. A test for survival suggested by Robert Thouless ("A test of survival", PSPR, 1948, 48, part 175, 253-263), in which a person encodes a message in a cipher of their own devising, unbreakable by rational means, with the intention of communicating after their death the key to the decipherment of the message. Cf. COMBINATION LOCK TEST FOR SURVIVAL. (D & W)

CIRCLE. A group of persons sitting with a medium, generally in a circle, possibly holding hands or with knees touching, or both, purportedly to establish a "psychic current" which can be used by the medium and/or discarnate entities for paranormal manifestations. A "Circle for Development" or "Development Circle" is one which meets for the purpose of cultivating mediumistic powers among the members, under the guidance of an acknowledged medium. A "home circle" is a circle which meets regularly at the home of the medium or of the sitters. (A)

CLAIRAUDIENCE. Paranormal information expressed as an auditory experience; it is generally considered a form or mode of clairvoyance. (F. clair, "clear", + audience, "hearing", ultimately der. L. clarus, "clear", + audientia, der. audire, "to hear")

CLAIRVOYANCE. Paranormal acquisition of information concerning an object or contemporary physical event; in contrast to telepathy, the information is assumed to derive direct from an external physical source, not from the mind of another person; not to be confused with the vulgar interpretation of the term as meaning "knowledge of the future" (for which see PRECOGNITION). Hence, "precognitive clairvoyance": the paranormal acquisition of information concerning a target not yet in existence and which will never be known sensorially to any mind. Hence also "clairvoyant": a person endowed with a special talent for clairvoyance; not to be confused with its colloquial usage meaning "a fortune-teller". (F. clair, "clear", + voyant, fm. voir, "to see", ultimately der. L. clarus, "clear-seeing", and videre, "to see")

CLAIRVOYANCE, PRECOGNITIVE. See under CLAIRVOYANCE.

CLOCK-CARD TEST. A test for ESP devised by G.W. Fisk and A.M.J. Mitchell ("ESP experiments with clock cards: a new technique with differential scoring",

JSPR, 1953, 37, 1-14) in which the cards used as targets show the position of one hand of a clock, there being altogether 12 possible target positions on a circular array.

CLOSED DECK. Description applied to a test procedure in which the target order for each run is generated not by independent random selection of successive targets, but by randomization of a fixed set of targets (e.g. 25 ESP cards).

Cf. OPEN DECK.

COGNITION, PARANORMAL. See EXTRASENSORY PERCEPTION.

COGNITIVE ERROR HYPOTHESIS. The theory that psi-missing is caused not by the subject's unconscious negative motivation, but rather by some error made in the decision-making process, such as consistently rejecting the initial impression in favour of a later, more vivid one, when in fact that first impression is correct. Cf. CONSISTENT MISSING.

COINCIDENCE. The occurrence of two or more events at one time, apparently by mere chance, yet perceived by an observer as possessing some striking and unlikely cognitive relationship. See also SYNCHRONICITY.

COLLECTIVE HALLUCINATION. See under HALLUCINATION.

COMBINATION LOCK TEST FOR SURVIVAL. A test for survival suggested by Ian Stevenson in which a person sets a combination lock to a combination known only to themselves with the intention of communicating after their death the numbers to which the lock must be set in order to be opened. Cf. CIPHER TEST FOR SURVIVAL. (D & W)

COMMUNICATOR. A personality purporting to be that of a deceased individual which communicates with the living, usually through a medium. (D & W)

COMPRESSIVE. Term borrowed from psychology by Betty Humphrey ("Success in ESP as related to form of response drawings: I. Clairvoyance experiments", J.P., 1946, 10, 78-106) to describe a particular sort of form-quality of freehand drawings, viz. those which are based on a fearful and unimaginative concept of space. There appears to be a fairly consistent tendency for subjects displaying compressiveness to score better on tests of GESP, and for expansive subjects to score better on tests of clairvoyance.

CONFIDENCE CALL. A response that the subject feels relatively certain is correct, and is so indicated before the response is compared with its target.

CONFIRMATORY EXPERIMENT. A formal, large-scale experimental attempt to confirm a finding or hypothesis suggested by earlier exploratory work. Cf. PILOT EXPERIMENT. (D & W)

CONFORMANCE BEHAVIOUR. Key concept in a theory put forward by Rex Stanford ("Toward reinterpreting psi events", JASPR, 1978, 72, 197-214); it refers to the reorganizing of a relatively random, or unordered, system so that its new state more readily subserves the "disposition" or goals of another relatively structured, or ordered, system; psi is seen as a special case of conformance behaviour.

CONFUSION MATRIX. A matrix in which all the different target symbols used in an ESP test are arranged on one axis, all the different possible responses on the other axis, the entries in the matrix representing the number of times a given response was called against a given target.

CONSISTENT MISSING. Term introduced by R. Cadoret and J.G. Pratt ("The consistent missing effect in ESP", J.P., 1950, 14, 244-256) to refer to the tendency of some subjects in a forced-choice test of ESP to respond consistently to a certain target with the same incorrect call, e.g. by consistently calling "circle" when the target is in fact "square"; it suggests a process akin to misrecognition in ordinary sense-perception.

CONTROL. A personality purporting to be that of a deceased individual, believed to take control of the medium's actions and speech during trance, and/or who habitually relays messages from the communicator to the sitter.

COOPERATOR. See TARGET PERSON.

CORONA DISCHARGE. See KIRLIAN PHOTOGRAPHY.

CORRECT MATCHINGS, THE METHOD OF. A technique for evaluating degree of similarity between target and response, particularly in the context of a free-response test: n targets are displayed with their n responses (in randomized order), and a blind judge is required to match, in pairwise fashion, each response with its intended target; the response is either correctly matched with its

own target (in which case it is called a "hit", or "correct matching"), or it is matched with a target for which another response was intended (in which case it is a "miss", or "incorrect matching"); the total number of correct matches may thus range from zero to n.

CORROBORATOR. A witness in a case of spontaneous psi who either observed the experient at the time of their experience, observed events connected with it, or received a report about it soon after its occurrence, thus being in a position to confirm in whole or in part the experient's own account. See also AUTHENTICATION. (D & W)

COVARIANCE EFFECT. Term used by Gardner Murphy and Ernest Taves to describe the finding that if ESP is evidenced in two tasks carried out concomitantly, high scores on one task tend to be accompanied by (i.e. "covary with") high scores on the other task, and the same for low scores mutatis mutandis. (D & W)

CR. See CRITICAL RATIO.

CR_d. See under CRITICAL RATIO.

CRITICAL RATIO (CR). A mathematical quantity used to decide whether or not the observed degree of deviation from chance in a psi test is significantly greater than the expected degree of random fluctuation about the average: it is obtained by dividing the observed deviation by the standard deviation; also called a "z-score". Hence, "critical ratio of the difference (CR_d): a CR representing the difference between the observed average scores of two samples of data, divided by the "standard deviation of the difference".

CROSS-CHECK. A comparison of the subject's calls with the targets used in some other part of the experiment for which those calls were not intended. (N)

CROSS-CORRESPONDENCE. A highly complex series of independent communications delivered to two or more geographically separate mediums such that the complete message is not clear until the separate fragments are pieced together into a meaningful whole.

CRYPTAESTHESIA. Term coined by Charles Richet to refer to extrasensory perception. (fm. Gk. kryptos, "hidden", + aisthēsia, "perceptive state", der. aisthanesthai, "to perceive")

CRYPTOMNESIA. A memory of an event or experience which has been forgotten by the conscious mind, and which may be recalled without the person recognizing it as a memory; sometimes invoked as a counterhypothesis to apparent paranormal awareness. (fm. Gk. kryptos, "hidden", + mnēsis, "memory")

CRYSTAL-GAZING. See SCRYING.

CUMBERLANDISM. See MUSCLE-READING.

CUTANEOUS PERCEPTION. See DERMO-OPTICAL PERCEPTION.

DEATHBED EXPERIENCE. An experience in which a dying person appears to be aware of the presence of deceased relatives or friends; alternatively, or in addition: an altered state of consciousness, such as exaltation, in the dying person. See also PEAK IN DARIEN CASE.

DECLINE EFFECT. The tendency for high scores in a test of psi to decrease, within either a run, a session, or over a longer period of time. Cf. INCLINE EFFECT; U-CURVE.

DEFENSE MECHANISM TEST (DMT). A projective test of personality developed by Ulf Kragh and inspired by Freudian theory, by means of which an individual's characteristic mechanisms for coping with anxiety can be determined: the subject is presented, by means of a tachistoscope, with a picture depicting a "hero" figure in the centre with a threatening face in the periphery; this peripheral stimulus appears to interfere with perception of the central figure, to the extent of the individual's underlying psychopathology; "perceptual defence", or "preconscious defence organization", as measured by this test, has been successfully related to scoring in a number of ESP tests.

DEGREES OF FREEDOM (df). In order to determine the probability, p, associated with a test of statistical significance, it is often necessary to know the relevant "degrees of freedom", i.e. the number of independent values that are free to vary in the series of numbers or scores contributing to the statistic in question.

DÉJÀ VU. The feeling or illusion of having previously experienced an event actually being encountered for the first time; also called "false memory", or "memory without recognition". (F. "already seen")

DERMO-OPTICAL PERCEPTION. The apparent ability to discriminate colour and brightness by means of touch. Also known as "skin vision", "cutaneous perception", "digital sight", "finger vision". (fm. Gk. derma, "skin", + optikos, "of sight", der. opsomai, "I shall see")

DEUTEROSCOPY. See SECOND SIGHT. (fm. Gk. deuteros, "second", + skopia, der. skopein, "to look at")

DEVIATION. The arithmetical quantity by which the observed score is above or below the mean chance expectation of a run, series or other unit of trials; can thus be either positive or negative.

df. See DEGREES OF FREEDOM.

DIAGNOSIS, PARANORMAL. The determination of the nature and circumstances of a diseased condition by means other than those recognized as possible by medical science. See also HEALING, PSYCHIC.

DIAMETRIC HYPOTHESIS. A hypothesis proposed by A.A. Foster ("Is ESP diametric?", J.P., 1940, 4, 325-328) according to which, when the components constituting a complex target situation are cognized by ESP, the cognition is achieved by a single act rather than by a step-by-step process.

DIFFERENTIAL EFFECT. In an experiment where the subjects are tested under two different procedural conditions, (i) the tendency of subjects who score above chance in one condition to score below chance in the other, and vice versa; (ii) the tendency of one condition to elicit psi-hitting from the group of subjects as a whole, the other condition, to elicit psi-missing; or (iii) the tendency of one condition to elicit non-significant positive scoring from one group, the other condition to elicit non-significant negative scoring, there being a significant difference between the two sets of scores.

DIGITAL SIGHT. See DERMO-OPTICAL PERCEPTION. (fm. L. digitus, "finger, toe")

DIRECT VOICE. A rare phenomenon associated with physical mediums in which a voice, whose source is ostensibly paranormal, is heard, often issuing from a small "trumpet" or megaphone which floats around the room.

DIRECT WRITING. The paranormal production of a written message.

DISCARNATE ENTITY. A disembodied being; the surviving personality of a deceased individual; a spirit. (fm. L. dis-, "away, apart", + caro (carnis), "flesh")

DISPLACEMENT. A form of ESP shown by consistently responding to a target at one or more remove, spatially or temporally, from the actual target designated for that trial. Hence, "backward displacement", in which the target extrasensorily cognized precedes the intended target by one, two or more steps (designated as -1, -2, etc.); "forward displacement", in which the target actually responded to occurs later than the intended target by one, two or more removes (designated as +1, +2, etc.). Hence also, "systematic displacement": term suggested by Michael Thalbourne ("A long-distance ESP drawing experiment between Austria and Iceland", Paper presented at the Third International S.P.R. Conference, Edinburgh, April 1979) to describe displacement where the "distance" between intended and cognized target is relatively constant (such as in -1 or +1 displacement), as opposed to "unsystematic displacement" (a phenomenon to be found only in free-response tests of ESP), where there is no discernible pattern of asynchrony, i.e., where the targets cognized display no systematic serial relation to their responses.

DIVINATION. The use of various practices, such as tea-leaf reading, palmistry, scrying, I Ching, Tarot cards, etc., to gain paranormal information.

DMT. See DEFENSE MECHANISM TEST.

DOCTRINAL COMPLIANCE. Concept introduced by Jan Ehrenwald to refer to the tendency of patients to recount dreams and other life-material which confirm the therapist's personal ideas and theories about psychotherapy; it is conjectured that telepathic leakage between patient and therapist plays a role in this phenomenon.

DOORWAY TEST. A test suggested by Charles Tart ("Concerning the scientific study of the human aura", JSPR, 1972, 46, 1-21) for studying a sensitive's alleged ability to perceive the aura of another person: a person supposedly endowed with a large, stable aura, positions themselves behind the edge of a doorway, such that no part of the physical body ever protrudes; a randomly-determined schedule of trials is arranged such that sometimes the target person stands distant from the doorway, and at other times, near to the doorway so that the aura will protrude several inches beyond it; the sensitive must say, at each trial, whether or not the person is near or far from the edge of the doorway,

and significant discrimination would be consistent with perception of an objective aura.

DOPPELGÄNGER. An apparitional double or counterpart of a living person. Cf. ASTRAL BODY. (German: "double walker")

DOUBLE. See DOPPELGÄNGER.

DOUBLE BLIND. See under BLIND.

DOWN THROUGH (DT). A technique for testing clairvoyance in which the percipient guesses the order, from top to bottom, of a pack of target cards before any are removed or compared with the call. Cf. UP THROUGH.

DOWSER. See under DOWSING.

DOWSING. A behavioural automatism in which a "dowsing rod" or "divining rod" (often a forked twig but sometimes a pendulum) is employed to locate subterranean water, ore, oil, etc. or other concealed items by following the direction in which the rod turns in the user's hands. Hence "dowser", a person who practises dowsing. Hence also "map dowsing", a form of dowsing in which the dowsing device indicates the location of the desired object on a map of the area, rather than in the geographical area itself. (origin unknown) (D & W)

DREAM, VERIDICAL. An apparently paranormal dream, inasmuch as some of the dream details give information about events normally unknowable to the experient.

DROP-IN COMMUNICATOR. Term coined by Ian Stevenson for a communicator who appears unbidden at a sitting, and entirely unknown to the medium, sitters, or anyone else present.

DT. See DOWN THROUGH.

DUAL ASPECT TARGET. See under MULTIPLE ASPECT TARGET.

DUAL TASK. Any test of psi where the scores obtained under two conditions are compared.

DUALIST HYPOTHESIS. The philosophical viewpoint that mental processes and physical processes are different in kind (and thus, that the former are not reducible to the latter); also, the view that psi is essentially non-physical, and that therefore the human mind, at least in part, is also non-physical and could in theory survive the death of the body, and possibly communicate with the

living in its discarnate form. See also SURVIVAL.

ECTOPLASM. Term coined by Charles Richet to describe the "exteriorized substance" allegedly emanating from the body of some physical mediums and out of which materializations are said to be formed. (fm. Gk. ektos, "outside", + plasma, "something formed or moulded")

EEG. See ELECTROENCEPHALOGRAPHY.

ELECTROENCEPHALOGRAPHY (EEG). A technique for amplifying and recording the fluctuations in electrical voltage in a living brain using electrodes attached to key positions on a person's head; the mechanical device itself is known as an "electroencephalograph", and the graphic record produced by it, an "electroencephalogram"; the EEG has proved to be particularly important for sleep-research (and thus also for research on dream-telepathy), where characteristic brain waves have been identified and related to the successive stages of sleep. (fm. Gk. enkephalos, "the brain", der. en + kephalē, "within" + "the head") (8,b)

ELECTRONIC VOICE PHENOMENA (EVP). Phenomena popularized by K. Raudive, consisting of sounds alleged to be the faint voices of deceased individuals, recorded on previously unused magnetic tapes.

ELECTROPHOTOGRAPHY. See KIRLIAN PHOTOGRAPHY.

ERROR PHENOMENON. Concept introduced by Ramakrishna Rao ("Spontaneous ESP in laboratory tests: the error phenomenon", JASPR, 1968, 62, 63-72) to describe the phenomenon wherein a procedural error made by the experimenter or by the subject either occurs as a result of psi or itself activates psi.

ESP. See EXTRASENSORY PERCEPTION.

ESP CARDS. A special deck of cards, first used by J.B. Rhine, for use in tests of ESP: a standard pack contains 25 cards, each portraying one of five symbols (circle, square, cross, star, waves). Also called Zener cards.

ESP PROJECTION. See OUT-OF-THE-BODY EXPERIENCE.

ETHERIC BODY. See ASTRAL BODY.

EVOKED POTENTIAL. A fluctuation in the electrical field of a part of the brain-cortex, produced when the person is subjected to any sort of stimulation, such as a flashing light or an auditory stimulus.

EVP. See ELECTRONIC VOICE PHENOMENA.

EXPANSIVE. Term borrowed from psychology by Betty Humphrey (1946) to describe a particular type of form-quality which may be found in freehand drawings, viz. form based on a bold, adventurous, imaginative, extraverted use of available space; not synonymous with "artistically skilful". There is a tendency for subjects showing expansiveness to obtain higher scores on tests of clairvoyance, compared with persons who are compressive, and lower scores on tests of GESP.

EXPECTANCY. See MEAN CHANCE EXPECTATION.

EXPECTATION. See MEAN CHANCE EXPECTATION.

EXPERIMENTER. The person who conceives and designs the experiment, and who can be considered the author and architect of its outcome; often, but not always, they conduct the testing of the subjects and other tasks, such as randomizing targets, ascertaining scores, and analyzing the data.

EXPERIMENTER EFFECT. An experimental outcome which results not from manipulation of the variable of interest per se, but rather from some aspect of the particular experimenter's behaviour, such as unconscious communication to the subjects as to what is expected of them, different ways of treating subjects, or possibly a psi-mediated effect working in accord with the experimenter's desire to confirm some hypothesis.

EXPLORATORY EXPERIMENT. See PILOT EXPERIMENT.

EXTENDED TELEPATHY. See SUPER-ESP HYPOTHESIS.

EXTRACHANCE. Not a result of chance alone. (fm. L. extra, "outside of")

EXTRASENSORY PERCEPTION (ESP). Paranormal cognition; acquisition of information about an external event, object or influence (mental or physical, past, present or future) otherwise than through any of the known sensory channels; term coined by J.B. Rhine to embrace such phenomena as telepathy, clairvoyance and precognition.

EYELESS SIGHT. The alleged ability to see when the eyes are covered; regarded by the Mesmerists as one of the paranormal concomitants of the deep trance induced by hypnosis, akin to clairvoyance; Jules Romains coined the expression "extra-retinal" or "paroptic" vision, and attempted to give it a physiological interpretation. See also DERMO-OPTICAL PERCEPTION. (B,a)

FIRE-IMMUNITY. See INCOMBUSTIBILITY; FIRE-WALKING.

FIRE-WALKING. The alleged phenomenon wherein a person is able to walk barefoot across a bed of red-hot coals without suffering any burns. See also INCOMBUSTIBILITY.

FOCUS. A living person, often a teenager, who appears to be the cause and focal point of the phenomena observed in a poltergeist outbreak; the poltergeist "agent" or "medium".

FOCUSING EFFECT. Term coined by J.G. Pratt for the tendency for success in a test of ESP to be concentrated upon particular individual target-cards more than upon others; also, a form of target-preference exhibited by the high-scoring subject Pavel Stepanek, who consistently favoured particular targets among a set of concealed objects which were placed in apparently similar pieces of cardboard. (Ryzl, M. & Pratt, J.G., "The focusing of ESP upon particular targets", J.P., 1963, 31, 192-197)

FORCED-CHOICE TEST. Any test of ESP in which the subject is required to guess a target which is one of a limited range of possibilities and known to them in advance; an example is a test using the ESP cards. Cf. FREE-RESPONSE TEST

FORCED MATCHING, THE METHOD OF. See CORRECT MATCHINGS, THE METHOD OF.

FORWARD DISPLACEMENT. See under DISPLACEMENT.

FREE-RESPONSE TEST. Any test of ESP in which the range of possible targets is relatively unlimited and unknown to the subject, thus permitting them to respond freely with any impressions that come to mind; an example is the drawing-reproduction technique, where a percipient guesses and draws the content of a drawing or other pictorial representation. Cf. FORCED-CHOICE TEST

GALVANIC SKIN RESPONSE (or REFLEX)(GSR). A decrease in the electrical resistance of the skin, in response to a psychophysiological stimulus; of interest to parapsychologists as a measure of emotional arousal in a percipient, which may be influenced by extrasensory stimuli. (named after the Italian physiologist Luigi Galvani, 1737-1798)

(M)

(γ)-TELEPATHY. Term coined by R.H. Thouless and B.P.Wiesner ("The Psi processes

in normal and "paranormal" psychology", PSPR, 1947, 48, part 174, 177-196) to refer to one of two possible kinds of telepathic process, in which the percipient acquires extrasensory information concerning the mind (or nervous system) of an agent, as opposed to the agent directly influencing the mind (or nervous system) of the percipient. See also (κ)-TELEPATHY; PSI-GAMMA; PSI-KAPPA

GANZFELD. Term referring to a special type of environment (or the technique for producing it) consisting of homogeneous, unpatterned sensory stimulation: audio-visual ganzfeld may be accomplished by placing halved ping-pong balls over each eye of the subject, with diffuse light projected onto them from an external source, together with the playing of unstructured noise into the ears, and generally with the person in a state of bodily comfort; the consequent deprivation of patterned sensory input is said to be conducive to introspection of inwardly-generated impressions, some of which may be extrasensory in origin. (German: "entire field")

GELLER EFFECT. The supposed ability to bend metal by paranormal means; named after the Israeli psychic Uri Geller, who was the first to claim the metal-bending ability. See also MINI-GELLER; PSYCHOKINESIS

GENERAL EXTRASENSORY PERCEPTION (GESP). A non-committal technical term used to refer to ESP in which the information may have been paranormally acquired from another person's mind (i.e., as telepathy), from a physical event or state of affairs (i.e., as clairvoyance), or from both sources; experimental parapsychologists rarely use the term "telepathy" because of the difficulty, in so-called tests of telepathy, of excluding the possible operation of clairvoyance.

GESP. See GENERAL EXTRASENSORY PERCEPTION.

GHOST. As popularly used, this term denotes only the apparition of a deceased person, and is not sufficiently precise for use in psychical research. (A)

GIFT OF TONGUES. See GLOSSOLALIA.

GLOSSOLALIA. Speaking in "tongues", that is, in a language which is either unknown to linguistic science, or completely fabricated; it usually occurs in a religious context or is attributed to religious inspiration, such as the Holy Spirit; not to be confused with xenoglossy. (fm. Gk. glōssa, "tongue, language", + lalia, "chat, gossip, talking", der. lalein, "to make an inarticulate sound")

GOAT. Originally, a term borrowed from a New Testament metaphor by Gertrude Schmeidler to describe a subject who rejects the possibility that ESP could occur under the conditions of the given experimental situation; this somewhat narrow meaning has been extended to refer also, or alternatively, to persons who do not believe in the existence of ESP in general (i.e. under any conditions!), or even to persons who obtain low scores on various projective, scalar or checklist measures of belief in and/or experience of different sorts of putative psi phenomena. Cf. SHEEP. See also SHEEP-GOAT EFFECT.

GREVILLE CORRECTION. A method developed by T.N.E. Greville for the statistical evaluation of data which consist of two or more sets of responses intended for one and the same set of targets. See also STACKING EFFECT

GSR. See GALVANIC SKIN RESPONSE (or REFLEX).

GUESS. The subject's overtly made call or response to the target in a test of ESP.

HALLUCINATION. An experience having the same phenomenological characteristics as a sense-perception, and which may lead the experient to suppose the presence of an external physical object as the cause of that experience, but where in fact there is no such object present. Hence "collective hallucination": a hallucination experienced simultaneously by two or more persons who are together. Hence also "reciprocal hallucination": hallucination, elements of which are shared by two or more persons out of sensory range of one another. Hence also "veridical hallucination": a hallucination which conveys correct information about an event or circumstance unknown to the experient. (D & W)

HALO. See AURA.

HAUNTING. The more or less regular occurrence of ostensibly paranormal phenomena associated with a particular locality (especially a building) and usually attributed to the activities of a discarnate entity; the phenomena may include apparitions, poltergeist disturbances, cold draughts, sounds of steps and voices, and various odours.

HEALING, PSYCHIC. Healing apparently brought about by such non-medical means as prayer, the "laying on of hands", immersion at a religious shrine, etc., and

inexplicable according to contemporary medical science; not to be confused with merely unconventional medicine; also, "absent healing": psychic healing brought about when the healer and patient are outside the sensory range of each other.

HIGH AIM. Psi consciously directed by the subject to produce hits. Cf. LOW AIM

HIGH-DICE TEST. A test for psychokinesis in which the subject attempts to influence a pair of dice to fall with the two uppermost faces totalling eight or more. Cf. LOW-DICE TEST

HIGH-SCORING SUBJECT. A person who consistently obtains extrachance results in psi experiments. (D & W)

HIGH VARIANCE. See under VARIANCE.

HIGH-VOLTAGE ELECTROPHOTOGRAPHY. See KIRLIAN PHOTOGRAPHY.

HIT. In general, a "correct" response in a test of psi: e.g. in a forced-choice test of ESP, a trial in which the call is the same as the target; in a free-response test, a response which displays a degree of correspondence to its target which exceeds a specified criterion (e.g. by being above a certain point on a dichotomized ordinal scale); in a test of psychokinesis, a trial in which the outcome of the target system (e.g. falling dice, or a random number generator) matches the aim of the subject.

HYPERAESTHESIA. A heightened acuteness of perceptual sensitivity; sometimes postulated as a counterhypothesis to instances of alleged ESP.
(fm. Gk. hyper, "beyond, overmuch", + aisthēsia, "perceptive state")

HYPNAGOGIC. Term referring to the transitional state of consciousness experienced while falling asleep, sometimes characterized by vivid imagery of varying bizarreness; also used to refer to the similar state of awareness experienced during the process of waking up. Cf. HYPNOPOMPIC. (fm. Gk. hypnos, "sleep", + agōgos, "leading")

HYPNOPOMPIC. Term referring to the transitional state of consciousness experienced upon wakening from sleep; the term "hypnagogic" is often used to refer to this state also. (fm. Gk. hypnos, "sleep", + pompos, "escort, guide")

HYPNOSIS. A sleep-like condition or state, which is accompanied by narrowing of the range of attention, characterized by marked susceptibility to suggestion and

which can be artificially induced. Hence, "hypnosis-at-a-distance": the induction of hypnosis in a person who is outside the sensory range of the hypnotist.

(modern L., der. Gk. hypnoun, "to put to sleep")

HYPNOTISM. The technique of inducing and exploiting hypnosis, by means of the suggestions and operations of the hypnotist, with whom the hypnotized person remains in rapport, responsive to their suggestions.

INCLINE EFFECT. The tendency for scores in a test of psi to increase over the duration of the experimental unit. Cf. DECLINE EFFECT; TERMINAL SALIENCE

INCOMBUSTIBILITY. The alleged ability to come into direct contact with fire or red-hot coals without being burned, as for example in fire-walking.

INCORPOREAL PERSONAL AGENCY (IPA). A discarnate entity which is entirely bodiless and a true agent.

INDUCTOR. See TOKEN-OBJECT.

INTERVENTION. Term used by Louisa Rhine to refer to the possibility that the recipient of a spontaneous precognitive impression can avert or avoid the foreseen event, or at least some aspects of it. See also INTERVENTION PARADOX

(D & W)

INTERVENTION PARADOX. If a person has a precognitive impression of an event E, and as a result of this information takes action such that E subsequently fails to occur, then the Intervention Paradox is the question of what it was that caused the precognitive impression in the first place, since the putative cause, viz. the future event E, never came into being. See also INTERVENTION

IPA. See INCORPOREAL PERSONAL AGENCY.

JUDGE. The person who awards a rating or rank-score to one or more responses produced (or targets used) in a free-response test of ESP, in accordance with the degree of correspondence obtaining between them and one or more targets (or responses); also, the person who, under "blind" conditions, attempts to match a set of targets with a set of responses. See also CORRECT MATCHINGS, THE METHOD OF; PREFERENTIAL MATCHING, THE METHOD OF.

K-IDEA. Term coined by Whately Carington (1945) to refer to an idea (other than that of the target) shared by agent and percipient, and serving to facilitate telepathic transmission of the idea of the target by means of the associative bond; the greater the number of K-ideas, the higher the probability of successful telepathic transmission. See also K-OBJECT; ASSOCIATION THEORY OF TELEPATHY

K-OBJECT. An object (often a photograph or personality description) associated with the agent and present with the percipient during an experiment in order to enhance the shared K-idea and thus facilitate telepathic transmission of the target by means of association. See also ASSOCIATION THEORY OF TELEPATHY (D & W)

(κ)-TELEPATHY. Term coined by R.H.Thouless and B.P. Wiesner (1947) to refer to one of two possible kinds of telepathic process, in which the agent causes, by psi-kappa, some change in the brain or motor part of the percipient's nervous system, as opposed to the percipient acquiring, by psi-gamma, information regarding the agent's mind or brain. See also (γ)-TELEPATHY; MOBIA; TELERGY

KEY CARDS. The "master" cards against which the subject in a test of ESP attempts to match the target cards.

KIRLIAN PHOTOGRAPHY. A type of high-voltage, high-frequency photography, developed in the Soviet Union by Semyon Davidovich Kirlian, which records the so-called "corona discharge" of an object, by causing ionization of the surrounding field; it is uncertain whether this corona is simply an artifact produced by the process itself, or whether it indicates the existence of hitherto unknown radiations or energy fields such as "bioplasma" or the psychic aura.

LATENT TELEPATHY. See under TELEPATHY.

LAYING-ON-OF-HANDS. See HEALING, PSYCHIC.

LEVITATION. The raising or suspension of persons or objects into the air without any apparent agency as required by known physical laws of motion and gravity. Hence, "autolevitation": the raising of the psychic's own body by their own paranormal power. (fm. Gk. autos, "self", + L. levitas, "lightness in weight")

LINGER EFFECT. A psychokinetic effect continuing to occur in a particular location after it is no longer being consciously evoked by the agent.

LINKAGE DESIGN. A design used in proxy sittings with mediums in which a chain of human links is interposed between the medium and the living source of information (usually the distant sitter) in order to render less likely the transfer of information from the latter to the medium by means of ESP; each participant in the experimental design is aware only of the link(s) adjacent to themselves in the chain. A similar design may be used in GESP experiments. (D & W)

LOW AIM. Psi consciously directed by the subject so as to produce misses. Cf. HIGH AIM

LOW-DICE TEST. A test of psychokinesis in which the subject aims to influence a pair of dice to fall with the two uppermost faces totalling six or less. Cf. HIGH-DICE TEST.

LOW VARIANCE. See under VARIANCE.

LUCID DREAM. A dream in which the dreamer is conscious of the fact that they are dreaming.

LUCIDITY. An early term for "clairvoyance", probably introduced into parapsychology by Charles Richet, but first found in the writings of the Mesmerists.

MACRO-PK. See under PSYCHOKINESIS.

MAIMONIDES BINARY TARGET POOL. A special series of 1,024 pictorial targets developed at the Maimonides Medical Centre, New York, by Charles Honorton ("Objective determination of information rate in psi tasks with pictorial stimuli", JASPR, 1975, 69, 353-359); each target represents a unique combination of elements from ten content categories (such as colour, activity, animals, nature features, etc.), such that each target constitutes ten independent binary (i.e. two-choice) trials; they can be used to provide an objective basis for quantifying the information content of the target, the subject's report, and the degree of correspondence between them.

MAJORITY VOTE TECHNIQUE. The so-called "repeated or multiple guessing technique" of ESP testing: the symbol called most frequently by a subject (or group of subjects) for a given target, is regarded as being the response to that target.

MAP DOWSING. See under DOWSING.

MATERIALIZATION. A phenomenon of physical mediumship in which living entities or inanimate objects become temporarily visible in apparently solid form, possibly shaped out of ectoplasm, by supposedly paranormal means.

MCE. See MEAN CHANCE EXPECTATION.

MEAN CHANCE EXPECTATION. The average number of hits, or the most likely score, to be expected in a test of psi on the null hypothesis that nothing apart from chance is involved in the production of the score.

MEAN VARIANCE (THEORETICAL). See under VARIANCE.

MEDITATION. A broad term embracing a number of techniques for achieving various altered states of awareness, with some of these altered states resulting in the ecstatic qualities of so-called "peak" experiences; most techniques are ways of learning to still the agitation of the mind so that more subtle and valuable aspects of Self and reality may be perceived; some techniques involve concentration, in which attention is focused on a particular object and restrained from wandering; others involve "expansiveness", where total attention is given to whatever spontaneously happens, with no attempt to control or focus attention. See also TRANSCENDENTAL MEDITATION.

MEDIUM, MENTAL. A person who regularly, and to some extent at will, receives impressions purporting to be communications from the dead, and who transmits these to the living, in oral or written form. See also COMMUNICATOR; CONTROL; TRANCE; MEDIUM, PHYSICAL

MEDIUM, PHYSICAL. A person who, usually while sitting with a group of other people, produces physical effects alleged to be paranormal and sometimes attributed to discarnate agency. See also APPORT; ECTOPLASM; MATERIALIZATION; LEVITATION (RSW)

MEDIUMSHIP, POLYGLOT. See XENOGLOSSY.

MENTAL CONTAGION. Expression used by René Warcollier to indicate the "spread" of impressions, with or without alteration, from mind to mind; under experimental conditions, the extrasensory reception of the same stimulus by a number of people simultaneously; implicated in crowd behaviour, for example.

MENTAL HEALING. See HEALING, PSYCHIC.

MENTAL OR BEHAVIOURAL INFLUENCE OF AN AGENT (MOBIA). Mental or behavioural effect induced in a percipient by an agent; term suggested by Rex Stanford

("An experimentally testable model for spontaneous psi events. II. Psychokinetic events", JASPR, 1974, 68, 321-356) to replace "active-agent telepathy", hypothesizing that the agent can play an active role in telepathy and that such "telepathy" is really a form of psychokinesis. Cf. KAPPA-(K)-TELEPATHY; TELERGY (Analogically formed fm. L. movere, "to induce movement", or mobilis, "moveable")

MENTAL PHENOMENA. Expression frequently used by Spiritualists to refer to those phenomena of mediumship that consist of alleged messages from the dead, communicated via the mind and body of the medium. See also MEDIUM, MENTAL; Cf. PHYSICAL PHENOMENA

MENTALISM. The practice of simulating telepathy, performed for the purpose of entertainment. Hence "mentalist", a person who engages in mentalism.

MESMERISM. Older term for "hypnotism", named after the Austrian physician Franz Anton Mesmer (1733-1815), who believed that it involved the transfer from operator to patient of a subtle fluid, force or energy known as "animal magnetism".

METAGNOME. French term for an especially talented percipient of paranormal impressions. (fm. Gk. meta, "after", + gnōmōn, "one that knows, interpreter")

METAPSYCHICS. Anglicization of a French term coined by Charles Richet as an alternative designation for the subject matter of parapsychology.

(fm. Gk. meta, "after", + psychikos, "of the soul, mental")

MICRO-PK. See under PSYCHOKINESIS.

MIDAS TOUCH IN REVERSE. Phrase coined by Ernest Taves and Laura Dale ("The Midas Touch in psychical research", JASPR, 1943, 37, 111-118) to describe the tendency of scores in parapsychological experiments in which experimenter and subject are aware that the ongoing results are in line with the experimenter's objectives in the first stages of the work, to drop thereafter to chance or below; the effect is attributed to both experimenter and subject, and appears to be a form of decline effect. (fm. Gk. legend: Midas was a Phrygian king who was given by Dionysus the power of turning into gold whatever he touched)

MIND READING. See TELEPATHY.

MINI-GELLER. A child or young person who can to some extent duplicate the metal-bending feats of Uri Geller, by apparently paranormal means. See also GELLER EFFECT

MINUS FUNCTION. An expression used by Jan Ehrenwald to describe the observation that psi phenomena are likely to occur during sleep and other "deficit states" of the organism, suggesting that psi is a compensatory extension of normal capacities. (D & W)

MISS. In general, an "incorrect" response in a test of psi, the opposite of a hit: in a forced-choice test of ESP, a trial in which the call is not the same as the target; in a free-response test, a response which displays a degree of correspondence to its target which falls short of a specified criterion (e.g. by being below a certain point on a dichotomized ordinal scale); in a test of PK, a trial in which the outcome of the target-system (e.g. a random number generator or falling dice) fails to match the aim of the subject.

MOBIA. See MENTAL OR BEHAVIOURAL INFLUENCE OF AN AGENT.

MULTIPLE-ASPECT TARGET. A target with two or more features each of which may or may not be guessed correctly by the percipient; an example is playing cards, which have three aspects, viz. colour, suit and number. If there are only two aspects, then the target may be referred to as a "dual aspect" one.

MULTIPLY-DETERMINED TARGET. Target involving more than one sense modality, for example sight combined with sound. (N)

MUSCLE-READING. A non-paranormal phenomenon which mimicks telepathy, in which a person is able, for example, to find a hidden object by means of physical contact with the person who knows its whereabouts, due to subtle muscular cues the latter provides unconsciously; also known as "Cumberlandism". (D&W)

NDE. See NEAR-DEATH EXPERIENCE.

NEAR-DEATH EXPERIENCE (NDE). Term applied to experiences undergone by persons who either seem to be at the point of death (or who are even formally declared dead) but then recover, or who narrowly escape death (as in a motor car accident) without being seriously injured; it has been suggested that there is, on coming close to death, a "core" NDE made up of certain common elements, such as a feeling of indescribable peace, a sense of being out of one's body, a movement into a dark void or down a tunnel, seeing a brilliant light, and entering into that light; panoramic memory (the "life review"), encountering an "unseen presence", being

greeted by deceased relatives or religious figures, may also be reported. See also DEATH-BED EXPERIENCE

NEWSPAPER TEST. A test for survival sometimes conducted at mediumistic sittings in which an attempt is made to exclude telepathy between medium and sitter by inducing the communicator to transmit a message referring to topics on specific pages of a newspaper that the medium could not normally have seen; this test can yield evidence of precognition when reference is made to newspapers not yet published. Cf. BOOK TEST (D & W)

NON-INTENTIONAL PSI. Term used by Rex Stanford to denote ESP-responses to need-relevant information which one is not consciously trying to obtain; the term is extended to encompass the non-intentional use of psychokinesis implicitly guided by extrasensory means. (D & W)

NULL HYPOTHESIS. The hypothesis provisionally held to be true whenever a statistical analysis of experimental data is undertaken, and generally asserting that the results can be explained as entirely due to chance factors; if the appropriate test of statistical significance produces results that are very improbable under the assumption that the null hypothesis is true ("very improbable" usually being defined as "having a probability of 1 in 20 or less, or odds of 19 to 1 or greater), the decision is made to reject the null hypothesis, and the findings are tentatively accepted as being due to causes specified in the experimental hypothesis. Sometimes symbolized as H_0. See also TYPE I (ALPHA) ERROR; TYPE II (BETA) ERROR

OBE. See OUT-OF-THE-BODY EXPERIENCE.

OBJECT-READING. See TOKEN-OBJECT TEST.

OBSERVATIONAL HYPOTHESIS. Term introduced by Sybo Schouten ("Testing some implications of a PK observational theory", EJP, 1977, 1, no.4, 21-31) to embrace a number of theories -- associated particularly with Helmut Schmidt and Evan Harris Walker -- which share the common features that they start from the conceptual tools of theoretical physics, and from the postulate that observation of the outcome of an experimental unit is a necessary though insufficient condition for psi

to influence the events leading to that outcome.

OBSESSION, SPIRIT. Supposed partial invasion of the mind of a living person by a discarnate entity, usually for the purpose of the latter's gratification. Cf. POSSESSION. (D&W)

OCCULT. Term referring to certain reputed sciences, such as magic, astrology, divination, etc., involving esoteric knowledge or employment of mysterious agencies; not to be confused with scientific parapsychology. (fm. L. occultus, "covered over, concealed")

OCCULTISM. The doctrine or study of the occult. Hence, "occultist", one who studies or practises occultism.

ONE-TAILED TEST. A test of statistical significance is said to be "one-tailed" if the direction of the descriptive statistic has been specified in advance as being positive (i.e. in the above-chance region), or, alternatively, has been specified as being negative (i.e. in the below-chance region); if the direction of the statistic is opposite to that predicted, then strictly speaking the result should be ignored, no matter how deviant, and the null hypothesis not be rejected. Cf. TWO-TAILED TEST

OOBE. See OUT-OF-THE-BODY EXPERIENCE.

OPEN DECK. Description applied to a procedure for generating a target order in which each successive target is chosen at random, independently of all the others; thus, for example, in the case of a standard deck of ESP cards whose target order is "open deck", each type of symbol is not necessarily represented an equal number of times (although over a sufficiently large number of trials, the proportion of times each of the n symbols occurs will converge upon 1/n). Cf. CLOSED DECK

OPTIONAL STOPPING HYPOTHESIS. In any ongoing series of trials in a test of psi, it is rather unlikely that at any one time the total number of hits will be exactly equal to mean chance expectation, even if only chance is operating; sometimes the total will exceed MCE, and it would be possible arbitrarily to stop an experiment at such a "favourable" point (this being termed "optional stopping"), and the optional stopping hypothesis has it that such a total, though significant, is spuriously so, that is, due entirely to chance. Sometimes incorrectly called "optimal stopping".

OSTENSIBLE. Adjective indicating that there is a possibility of paranormal causation being present, but that such a factor has not been conclusively demonstrated. (A)

OUIJA BOARD. A device consisting of a small board on legs, at whose apex is a pointer, and which rests on a larger board marked with words, alphabetical letters and numerals; by moving over the larger board and stopping at the various markings, while the fingers of one or more persons rest lightly upon it, the pointer is employed to give messages, answers, etc. Sometimes a glass tumbler is substituted for the moving board. See also PLANCHETTE (F. oui + German ja, both meaning "yes")

OUT-OF-THE-BODY EXPERIENCE (OBE). An experience, either spontaneous or induced, in which one's centre of consciousness seems to be in a spatial location outside of one's physical body; Celia Green distinguishes two types of "ecsomatic" or out-of-the-body experiences: the "parasomatic", where the person appears to themselves to possess a duplicate body, sometimes connected to the physical body by a "silver cord"; and the "asomatic", where they feel themselves to be disembodied; in either case, many experients claim to perceive their physical body lying inert, to see and hear people while remaining unperceived themselves, and to perceive objects and events normally beyond the range of their physical senses; of special interest to parapsychologists on account of its alleged connection with clairvoyance, and to students of survival as providing an example of what disembodied existence could be like. Note also the type of OBE known as "ESP projection", in which the person is seen by other people at a distant point and/or afterwards reports a veridical description of what he or she observed at that point. The OBE is also known as "astral projection" or "astral travel", and is similar to travelling clairvoyance. See also AUTOSCOPY; ASTRAL BODY (Dfw), (Bja)

P (PROBABILITY). The a priori (i.e. chance) probability of success per trial; expressed as the reciprocal of the number of different choices or targets when each kind of target has an equal probability of occurrence; for example, in a test with five equiprobable targets, P = 1/5, or one success in five trials. Cf. p p (probability). The fraction of times (i.e. the likelihood) that, in a large

number of repetitions of the experiment, an outcome as extreme or more so than the one actually observed will be obtained merely by chance; if the p-value is less than or equal to a certain value, viz. alpha, it is usual to reject the null hypothesis and to consider the result significant. Cf. P (N.B.: the definitions attached to P and p are often found interchanged, and the usage is inconsistent.)

PARAGNOSIA. Term used in the Netherlands for ESP. (fm. Gk. para, "beside, beyond", + gnōsis, "knowledge") See also PARAGNOST

PARAGNOST. Term used in the Netherlands for someone who displays special ability for paragnosia; a sensitive. (fm. Gk. para, + gnōstēr, "one that knows the truth of a thing")

PARAMNESIA. A distortion of memory or recognition; term often used synonymously with its best known form, déjà vu. (fm. Gk. para, + mnēsis, "memory") (D & W)

PARANORMAL. Term applied to any phenomenon which in one or more respects exceeds the limits of what is deemed physically possible on current scientific assumptions; often used as a synonym for "psychic", "parapsychological", "attributable to psi", or even "miraculous" (though shorn of religious overtones). (fm. Gk. para, "beside, beyond", + normal)

PARANORMAL HEALING. See HEALING, PSYCHIC.

PARAPHYSICAL. Pertaining to paraphysics; synonym for "psychokinetic".

PARAPHYSICS. That part of parapsychology which deals with the physical aspects of psi phenomena. Hence "paraphysicist", a scientist specializing in paraphysics.

PARAPSYCHIC. Term used by J.B. Rhine to refer to a high-scoring subject in parapsychological research.

PARAPSYCHICAL. Attributable to the operation of psi; synonym for "paranormal".

PARAPSYCHOLOGY. Term coined by Max Dessoir and adopted by J.B. Rhine to refer to the scientific study of paranormal phenomena, i.e. psi; except in Britain, it has largely superseded the older expression "psychical research". Hence, "parapsychological", and "parapsychologist", a scientist who specializes in parapsychology. (fm. Gk. para, "beside, beyond", + psychology, der. Gk. psychē, "soul, mind", + logos, "rational discussion")

PASSAGE OF MATTER THROUGH MATTER. See APPORT; TELEPORTATION.

PEAK IN DARIEN CASE. Term introduced by Frances Cobbe to indicate a special type

of death-bed experience in which the dying person seems to be aware of the presence of a deceased relative or friend of whose prior death they could not have been normally aware. (D & W)

PENDULUM. A device, such as a weight on the end of a thread, which is used in dowsing and a number of other divinatory practices.

PERCIPIENT. The person who experiences or "receives" an extrasensory influence or impression; also, one who is tested for ESP ability. Cf. AGENT; SUBJECT (JR)
(fm. L. percipiens, der. percipere, "to receive, understand")

PERSONA THEORY. Hypothesis proposed by Hornell Hart postulating that apparitions of the dead and mediumistic communicators are temporary re-creations of the personality-structures (i.e. "personas") of the deceased by the unconscious dramatizing powers of the percipient or the medium; however, they may (according to Hart) sometimes derive ingredients from actual discarnate entities.
(fm. L. persona, "actor's mask, character acted") (D & W)

PHANTASM. Any hallucinatory sensory impression, whatever sense may happen to be affected. See also APPARITION; HALLUCINATION (fm. Gk. phantasma, "appearance, image, phantom") (N)

PHOTOGRAPHY, PARANORMAL. The paranormal production of images on photographic film; also known as "thoughtography", a term first used to describe the experiments of Fukurai but adopted by Jule Eisenbud to describe the phenomena produced by Ted Serios, as if mental images were "projected" onto the film. See also SPIRIT PHOTOGRAPHY

PHYSICAL PHENOMENA. Term usually associated with Spiritualism and referring to those phenomena of mediumship which involve the production of ostensibly paranormal physical effects. See also MEDIUM, PHYSICAL; Cf. MENTAL PHENOMENA (D&W)

PILOT EXPERIMENT. A preliminary or exploratory test carried out for such purposes as the selection of suitable subjects, determining workable experimental conditions, and/or developing tentative hypotheses for testing in later full-scale investigations. Cf. CONFIRMATORY EXPERIMENT (D & W)

PK. See PSYCHOKINESIS.

PK PLACEMENT TEST. See PLACEMENT TEST.

PLACEMENT TEST. A technique for testing psychokinesis in which the aim of the subject is to try to influence falling objects to come to rest in a designated

area of the throwing surface. (JP)

PLANCHETTE. A small board supported by two castors and a writing-instrument positioned vertically, said to write messages without conscious effort by persons whose fingers rest lightly on the board. See also AUTOMATIC WRITING; OUIJA BOARD (fm. F., diminutive of planche, "planch")

PLETHYSMOGRAPH. An instrument used for determining variations in the size of an organ or of its light-transmitting characteristics, and hence variations in the amount of blood it contains; of use to parapsychologists as a measurer of emotional arousal in a percipient, which may be influenced by extrasensory stimuli. (N)

(fm. Gk. plēthuein, "to swell, increase")

PMIR. See PSI-MEDIATED INSTRUMENTAL RESPONSE.

PMS. See POST-MORTEM SURVIVAL.

POLTERGEIST. A disturbance characterized by bizarre physical effects of paranormal origin, suggesting mischievous or destructive intent: these phenomena include the unexplained movement or breakage of objects, loud raps, the setting of fires, and occasionally personal injury to people; in contrast to a haunting, the phenomena often seem to depend upon the presence of a particular individual (the "focus"), and apparitions are rarely seen. (German, literally "noise ghost")

POLTERGEIST MEDIUM. See FOCUS.

POLYGLOT MEDIUMSHIP. See XENOGLOSSY. (fm. Gk. polyglōttos, "many-tongued")

POSITION EFFECT. The tendency of scores in tests of psi to vary systematically according to the position of the trial on the record sheet. See also QUARTER DISTRIBUTION. Cf. SERIAL-POSITION EFFECT

POSSESSION. The complete control, by an ostensible discarnate entity, of the body of a living person. Cf. OBSESSION

POSTCOGNITION. See RETROCOGNITION. (fm. L. post, "after", + cognitio, "a getting to know")

POST-MORTEM COMMUNICATION. A communication or message purportedly from a deceased to a living person, usually delivered through a medium.

POST-MORTEM SURVIVAL (PMS). See SURVIVAL. (fm. L. post mortem, literally, "after death")

PQ. See PSI QUOTIENT.

PRATT-BIRGE METHOD. A method developed by J.G. Pratt and W.R. Birge ("Appraising verbal test material in parapsychology", JP, 1948, 12, 236-256) in which the Greville Correction is used in statistically evaluating the accuracy of data from free-response tests, especially the content of mediumistic communications. (D & W)

PRECOGNITION. A form of ESP in which the target is some future event that cannot be deduced from normally known data in the present. (fm. L. prae-, "prior to", + cognitio, "a getting to know")

PRECOGNITIVE CLAIRVOYANCE. See under CLAIRVOYANCE.

PRECOGNITIVE TELEPATHY. See under TELEPATHY.

PREFERENTIAL EFFECT. Term coined by Ramakrishna Rao ("The preferential effect in ESP", JP, 1962, 26, 252-259) to refer to a subclass of the differential effect in which the contrasting condition associated with positive scoring in the test for psi is the one for which the subject had an expressed preference.

PREFERENTIAL MATCHING, THE METHOD OF. A technique for evaluating the degree of correspondence between targets and responses in a free-response test of ESP: a judge examines a set of n comparison stimuli, and ranks them against one master stimulus in accordance with their degree of similarity to it; the set of comparison stimuli may consist of the targets used in the test (all of them, a subset, or the actual target plus a number of unused "control" targets), and be ranked against one response; or the set of comparison stimuli may be some or all of the responses produced in the test, and be ranked against one target; in either case, a rank-score of 1 is awarded to the comparison stimulus judged to correspond most highly to the master, a score of 2 to that stimulus next highest in correspondence, and so on to a score of n for the least similar.

PREMONITION. A feeling or impression that something is about to happen, especially something ominous or dire. See PRECOGNITION (fm. L. prae-, "prior to", + monitio, "warning")

PREVISION. See PRECOGNITION. (fm. L. prae-, + visio, der. videre, "to see")

PROBABILITY. See P; p

PROXY SITTER. A person who is present at a séance in the capacity of representing an absent subject, about whom they know nothing. See also PROXY SITTING.

(fm. Middle English prokecye, contraction of procuracy)

PROXY SITTING. A sitting at which the person desiring to receive a communication via a medium absents themselves from the actual sitting and is represented by another person, called a "proxy sitter".

Ψ. See PSI.

PSI (Ψ). A general blanket term, proposed by B.P. Wiesner and seconded by R.H. Thouless ("The present position of experimental research into telepathy and related phenomena", PSPR, 1942, 47, part 166, 1-19), and used either as a noun or adjective to identify paranormal processes and paranormal causation; the two main categories of psi are psi-gamma (paranormal cognition; ESP) and psi-kappa (paranormal action; PK), although the purpose of the term is to suggest that they might simply be different aspects of a single process, rather than distinct and essentially different processes. (Gk., psi, twenty-third letter of the Greek alphabet)

ψ-COMPONENT. See PSI-COMPONENT.

PSI-COMPONENT (ψ-COMPONENT). A concept introduced by C.D. Broad (Lectures on psychical research", New York: Humanities Press, 1962), on the assumption that it is logically impossible to be a person without any kind of body, and therefore inconceivable that a person might exist posthumously in a non-embodied state: Broad postulates that every human being is some kind of intimate "compound" of two constituents, one being their ordinary everyday body, and the other a "psi-component" (or "psychic factor"); the latter is assumed to have some of the quasi-physical properties of the traditional ghost, which carries some part at least of the dispositional basis of a person's personality, and is associated with a living brain and nervous system, and capable of persisting after the destruction of the physical body; apparently, not to be confused with "mind" or "astral body".

PSI-CONDUCIVE. Favourable to, or facilitative of, the occurrence of psi, whether it be manifested as psi-hitting or psi-missing. Cf. PSI-INHIBITORY

PSI-DEXTERITY. Term introduced by E.A.G. Knowles ("Psi dexterity in a mixing experiment", in J.B. Rhine and R. Brier (Eds.) Parapsychology Today, New York: Castle, 1968, pp. 29-34) to refer to paranormal success in an experiment where it is not possible to attribute conclusively the process to ESP or PK, but which may

be the result of either or both operating, especially in conjunction with the exertion of some physical, manipulative effort on the part of the subject in a situation where psi is necessary to bring about significant results.

PSI-DIFFERENTIAL EFFECT. See DIFFERENTIAL EFFECT.

Ψ_ϵ. See PSI-EPSILON.

PSI-EPSILON (Ψ_ϵ). Term coined by R.H. Thouless and B.P. Wiesner (1947) to refer to the form-producing activity of the soul, or Shin, normally manifested endosomatically as bodily growth and regeneration, but also "paranormally" manifested exosomatically as materialization and psychic healing; possibly the same process as psi-kappa. (Gk. epsilon, fifth letter of the Greek alphabet)

PSI-FIELD HYPOTHESIS. Theory proposed by William Roll ("The psi field", Proceedings of the Parapsychological Association, no. 1, 1957-1964, Durham, North Carolina: Parapsychological Association, 1966, pp. 32-65) in which it is postulated that physical events associated with physical objects produce changes in the "psi fields" surrounding those objects (a psi field being analogous to an electromagnetic or gravitational field), and that these "psi traces" are communicated via intermediary psi fields to the psi field of a percipient, where it interacts with their brain to produce an instance of ESP.

Ψ_γ. See PSI-GAMMA.

PSI-GAMMA (Ψ_γ). Term introduced by R.H. Thouless and B.P. Wiesner ("The psi processes in normal and "paranormal" psychology", PSPR, 48, 1947, part 174, pp. 177-196) as being more neutral than "ESP", since "perception" or "cognition" implies an awareness of a fact, which may or may not be present in a given instance. See also PSI. Cf. PSI-KAPPA (Gk. gamma, third letter of the Greek alphabet)

ψH. See PSI-HITTING.

PSI-HITTING. The use of psi in such a way that the target at which the subject is aiming is "hit" (i.e. correctly responded to, in a test of ESP, or influenced, in a test of PK), more frequently than would be expected if only chance were operating; the term is also sometimes used, misleadingly, to refer merely to non-significant positive scoring. Hence, "psi-hitter", a subject who exhibits a tendency to psi-hit. Cf. PSI-MISSING (Abbreviated to ψH by James Carpenter.)

PSI-INHIBITORY. Unfavourable to, or suppressive of, the occurrence of psi,

whether it be manifested as psi-hitting or psi-missing. Cf. PSI-CONDUCIVE

Ψ_K. See PSI-KAPPA.

PSI-KAPPA (Ψ_K) Term introduced by R.H. Thouless and B.P. Wiesner (1947) as an alternative to "psychokinesis", to distinguish the motor aspect of psi from its cognitive aspect, while not suggesting that the latter is any different in kind from the former. Cf. PSI-GAMMA; PSI-EPSILON (Gk. kappa, tenth letter of the Greek alphabet)

ΨM. See PSI-MISSING.

PSI-MASKING. The operation of psi in such a way that it covers up evidence of its own manifestation, as for example in a differential effect, decline effect, or cancellation effect.

PSI-MEDIATED. Caused by some aspect of psi.

PSI-MEDIATED INSTRUMENTAL RESPONSE (PMIR). A model proposed by Rex Stanford ("An experimentally testable model for spontaneous psi events. I. Extrasensory events", JASPR, 1974, 68, 34-57), in which it is proposed that an organism non-intentionally uses psi to scan its environment for need-relevant objects or events or for information crucially related thereto, and that when such information is obtained, the organism tends to make responses which are instrumental in satisfying its needs in relation to the particular object or event in question.

PSI-MEDIATING VEHICLE. Any symbol, image or object which facilitates the entry into conscious awareness of extrasensory information.

PSI-MISSING. The use of psi in such a way that the target at which the subject is aiming is "missed" (i.e. responded to incorrectly, in a test of ESP, or influenced in a direction contrary to aim, in a test of PK) more frequently than would be expected if only chance were operating; the term is also sometimes used, misleadingly, to refer simply to non-significant negative scoring. Hence, "psi-misser", a subject who displays a tendency to psi-miss. Cf. PSI-HITTING. (Abbreviated to ΨM by James Carpenter)

PSI PHENOMENON. Any event which results from, or is an instance of, the operation of psi; examples are the forms of ESP, and psychokinesis.

PSI QUOTIENT (PQ). A statistic introduced by Helmut Schmidt which is said to be

a measure of "psi efficiency" in a given test performance; it is obtained by squaring the critical ratio, dividing by the number of trials, and multiplying the result by 1,000. (fm. L. quotiens, "how many times?")

PSI-TRAILING. An animal's following its owner into wholly unfamiliar territory under conditions that preclude the use of a sensory trail (e.g. by scent); that is, "homing by psychic means".

PSYCHIC(AL). As a noun, "psychic" refers to an individual who possesses psi ability of some kind, and is thus a synonym for "sensitive"; as an adjective, it is applied to paranormal events, abilities, research, etc., and thus means "concerning psi, parapsychological". (fm. Gk. psychikos, "of the soul, mental")

PSYCHIC ETHER. A hypothetical substantial medium which propagates "thought waves", records impressions of events, and produces corresponding patterns. (N)
(fm. Gk. aithēr, "the upper, purer air; heaven")

PSYCHIC HEALING. See HEALING, PSYCHIC.

PSYCHIC PATHOLOGY OF EVERYDAY LIFE. Expression used by Jule Eisenbud to suggest that slips-of-the-tongue, memory lapses, apparently accidental physical events, and other anomalous occurrences become meaningful when viewed as instances of psi-determined or psi-conditioned behaviour. (D & W)

PSYCHIC PHOTOGRAPHY. See PHOTOGRAPHY, PARANORMAL.

PSYCHIC SHUFFLE. A test of psi devised by J.B. Rhine and colleagues (Rhine, J.B., Smith, B.M. & Woodruff, J.L., "Experiments bearing on the precognition hypothesis: II. The role of ESP in the shuffling of cards", JP, 1938, 2, 119-131) in which the subject is required to shuffle a deck of cards in such a way that the resulting order corresponds to that of a specified target-sequence.

PSYCHIC SURGERY. A form of psychic healing practised particularly in the Philippines, in which diseased tissues are allegedly removed without the use of surgical instruments, and bleeding, infection etc. are inhibited paranormally. The term is also used of surgery in which the surgeon operates while in a trance, as performed by J. Arigo and other Brazilian exponents of this practice, usually using unsterilized knives as scalpels. Hence, "psychic surgeon", a practitioner of this supposed art.

PSYCHICAL RESEARCH. The original term for "parapsychology", still widely used

in Britain; hence, "psychical researcher", a person who engages in psychical research.

PSYCHOBOLY. Term coined by A. Tanagras to denote psychokinesis, especially when used for malevolent purposes, as in alleged "evil eye" phenomena. (fm. Gk. psychē, + bolē, "a throw, stroke, wound of a missile") (D&W)

PSYCHOENERGETICS. Term used for "parapsychology" in the Soviet Union. (fm. Gk. psychē, + energētikos, "active", der. energein, "to work")

PSYCHOKINESIS (PK). Paranormal action; term introduced by J.B. Rhine to refer to the direct influence of mind on a physical system without the mediation of any known physical energy. Hence, "micro-PK": psychokinesis which has as its target a quantum-mechanical system; "macro-PK": psychokinesis which has, as its target, systems larger than quantum-mechanical processes, such as micro-organisms and macroscopic objects the like of dice; both terms coined by Brian Millar. See also PSI-KAPPA; RETRO-ACTIVE PK (fm. Gk. psychē, + kinēsis, "a moving, disturbance", der. kinein, "to set in motion"; makros, "long"; mikros, "small")

PSYCHOMETRIC OBJECT. See TOKEN OBJECT.

PSYCHOMETRIST. See TOKEN-OBJECT READER.

PSYCHOMETRY. See TOKEN-OBJECT TEST.

PSYCHON THEORY OF MIND. A theory proposed by Whately Carington (Telepathy. An outline of its facts, theory, and implications, Methuen, London, 1945) in which mind is viewed as consisting solely of "psychons" (images and/or sensa) grouped by means of associative links; he applies the theory to survival, suggesting that what survives death are the psychon systems resulting from earlier sensory stimuli together with others linked with them, especially through telepathic interaction with other minds (i.e. other psychon systems). (fm. Gk. psychē, + -on, suffix used in scientific terms to indicate "particle" or "entity", as in proton, ion, neutron) (D & W)

PSYCHOTRONICS. Czech term for "parapsychology", but embracing certain phenomena that are not now generally accepted as parapsychological. (D&W)

PURE CLAIRVOYANCE TECHNIQUE. A method of testing for clairvoyance in which a machine generates the order of the targets and scores the responses but records

only the total scores, thus ruling out telepathy (both contemporaneous and precognitive) as an explanation for extrachance results. (D & W)

PURE TELEPATHY TECHNIQUE. A method of testing for telepathy in which no physical record is ever made of the targets (which are instead only thought about by the agent), thus ruling out clairvoyance or precognition as an explanation for extrachance results.

PYGMALION HYPOTHESIS. Hypothesis proposed by Jule Eisenbud concerning survival: it states that while apparently surviving personalities may be living entities in certain restricted senses, they are not surviving portions of the deceased, but rather re-creations by the living of certain aspects of the deceased; once organized, such entities may have varying degrees of autonomy, but they continue to exist only for as long as their life is sustained by the living. (fm. Gk. legend: Pygmalion was a sculptor and king of Cyprus who fell in love with an ivory statue which he had carved and which, in answer to his prayer, Aphrodite endowed with life.) (D & W)

QD. See QUARTER DISTRIBUTION.

QUALITATIVE EXPERIMENT. Any test for psi which uses target material such that no definite probability of occurrence may be attached to the particular response items made; statistical evaluation must therefore proceed in an indirect fashion, by assigning a probability to the matching-performance of a judge. See also FREE-RESPONSE TEST (Ultimately der. L. qualis, "what kind of?")

QUANTITATIVE EXPERIMENT. Any test for psi which uses target material which has specific prescribed probability values, and which therefore allows direct statistical evaluation of the results obtained. See also FORCED-CHOICE TEST

(Ultimately der. L. quantus, "how great, how much?") (D & W)

QUANTUM MECHANICS. A branch of physics concerning the emission or absorption of energy by atoms and molecules or subatomic particles, the process taking place by steps, with each step being the emission or absorption of an amount of energy called the "quantum". (N)

QUARTER DISTRIBUTION (QD). A distribution of hits found not infrequently in the record page used in a test of psychokinesis, showing a decline in scoring both vertically down the page (i.e. during the course of a run), and horizontally

across the page (i.e. from run to run), thus resulting in the upper left-hand quarter of the page having a higher score than the lower right-hand quarter.

RADIAESTHESIA. A purported sensitivity to distant substances, said to be possessed by apparatus used for divining. (fm. L. radius, "rod, ray, beam of light", + Gk. aisthēsia, "perceptive state")

RANDOM BEHAVIOUR TRIAL. Expression introduced by Remy Chauvin and used predominantly in relation to research on anpsi to describe those trials on which the animal reacts, for no apparent reason, in a way which cannot be predicted on the basis of its habits, environmental stimulation, etc.

RANDOM EVENT GENERATOR. See RANDOM NUMBER GENERATOR.

RANDOM NUMBER GENERATOR (RNG). An electronic apparatus incorporating a randomizing element capable of generating a random sequence of outputs; used in automated tests of psi for generating target sequences, and in tests of psychokinesis may itself be the target system which the subject is required to influence, viz. by "biassing" the particular number or event output; a "binary" RNG has two equiprobable outputs; the term "RNG" is increasingly used to refer to any system which produces naturally random outputs, such as bouncing dice, radioactive decay, or even the brain.

RANDOM ORDER. A sequence of events which exhibits no trends or regularities that would permit prediction of a later event from an earlier one in the series.

RANDOMIZER. The person who prepares a random order of targets for an experiment on psi.

RAUDIVE VOICE PHENOMENA. See ELECTRONIC VOICE PHENOMENA.

READING. The statements made by a sensitive in the course of their attempt to obtain paranormal information or "messages". See also AURA-READING; TOKEN-OBJECT TEST

REBIRTH. Term generally used as a synonym for "reincarnation"; however, some Buddhists distinguish between "rebirth", as continuation from one life to another of personality only, and "reincarnation" as continuation of identity.

RECEIVER. Less technical expression than "percipient", used to indicate the

subject designated as the "recipient" of telepathic information. Cf. SENDER

RECIPROCAL HALLUCINATION. See under HALLUCINATION.

RECITATIVE XENOGLOSSY. See under XENOGLOSSY.

RECURRENT SPONTANEOUS PSYCHOKINESIS (RSPK). Expression coined by William Roll to refer to paranormal physical effects which occur repeatedly over a period of time, and especially used as a neutral descriptor of poltergeist disturbances. See also PSYCHOKINESIS

REG. See RANDOM NUMBER GENERATOR.

REINCARNATION. A form of survival in which the human soul, or some aspects of self, is, after the death of the body, reborn into a new body, this processes being repeated throughout many lives. See also REBIRTH (fm. L. re-, "again", + in, "into", + caro (carnis), "flesh")

REINFORCEMENT EFFECT. A phenomenon of displacement in which there is a higher displacement score when the target card lies between two cards bearing the same symbol than when it lies between two cards bearing different symbols.

RELEASE-OF-EFFORT EFFECT. See LINGER EFFECT.

REMOTE VIEWING. A neutral term for general extrasensory perception introduced by Russell Targ and Harold Puthoff ("Information transmission under conditions of sensory shielding", Nature, 1974, 251, 602-607), especially in the context of an experimental design wherein a percipient attempts to describe the surroundings of a geographically distant agent.

REPEATABLE. As applied to experimental findings, "repeatable" phenomena are those which can be produced by any competent independent investigator by following a particular standardized procedure. See also REPEATABILITY, THE PROBLEM OF; REPLICATE

REPEATABILITY, THE PROBLEM OF. The problem that parapsychological effects cannot be produced "on demand", i.e. that they lack the characteristic of being adequately "repeatable".

REPEATED-GUESSING TECHNIQUE. See MAJORITY-VOTE TECHNIQUE.

REPLICATE. As a transitive verb, "to replicate" means to obtain the same findings when a sufficiently similar experimental procedure is followed designed to yield

that effect; as an intransitive verb, "to replicate" means "to occur again" under a sufficiently similar experimental procedure.

REPLICATION. (i) an experiment which is designed to yield the same findings as a previous experiment or series thereof; Carl Sargent distinguishes between "concrete" and "conceptual" replication: an experiment which precisely duplicates the essential conditions of its predecessor is a concrete, or exact, replication, while an experiment which resembles its predecessor only to the extent that it tests the same hypothesis or theoretical construct is a conceptual replication; (ii) the event of conducting a "replication" study; (iii) more strictly, an experiment which does in fact "replicate" the findings of its predecessor(s); in this sense an experiment which attempts to replicate previous findings but fails is only an "attempted replication".

RESPONSE. In general, any overt piece of behaviour that occurs as a result of a stimulus; in the context of an ESP experiment, it often refers to the guess or call that the subject makes in attempting to identify the target.

RESPONSE BIAS. The tendency to respond or behave in predictable, non-random ways.

RESPONSE-BIAS HYPOTHESIS. Theory proposed by Rex Stanford ("Response bias and the correctness of ESP test responses", JP, 31, 1967, 280-289) which postulates that responses made which are counter to the organism's response bias are particularly likely to be correct (i.e. "hits"), on the grounds that (a) the degree of internal feeling that a signal or target is present (rather than absent) is correlated with the actual presence of that signal, and (b) making a counter-bias response requires an unusually strong sense of the presence of the related signal.

RESPONSIVE XENOGLOSSY. See under XENOGLOSSY.

RETROACTIVE CAUSATION. A supposed instance of causality in which a necessary condition for the occurrence of event A is the occurrence of a later event B.

(fm. L. retro, "backward, behind")

RETROACTIVE PK. Psychokinesis occurring in such a way as to be an instance of retroactive causation; to say that event A was caused by retroactive PK is to say that A would not have happened in the way that it did had it not been for a later PK effort exerted so as to influence it. Sometimes abbreviated to "retro-PK";

also referred to as "backward PK" or "time-displaced PK".

RETROCOGNITION. A form of ESP in which the target is some past event which could not have been learned or inferred by normal means. (fm. L. retro, "backward, behind", + cognitio, "a getting to know")

RETRO-PK. See RETROACTIVE PK.

REVERSAL EFFECT. Term used to refer to the finding in replication experiments of results in a direction opposite to those found in the original study.

RNG. See RANDOM NUMBER GENERATOR.

RSPK. See RECURRENT SPONTANEOUS PSYCHOKINESIS.

RUN. A fixed group of successive trials in a test of psi; it may be of any predetermined length, but usually consists of 25 calls of an order of target cards in a test of ESP, or of 24 throws of one or more dice in a test of PK; usually there is a pause between each run; the run is often the unit into which the session is subdivided for purposes of analysis.

RUN-SCORE VARIANCE. See under VARIANCE.

SALIENCE RATIO. A number representing a test for "terminal salience", in which the scoring rate in the end segments of the run (or in the end trials of the segment) is compared with the rate of scoring in the middle segments (or trials). See also U-CURVE (der. L. salire, "to leap")

SCHMIDT MACHINE. Term used to refer to a random number generator which makes use of radioactivity as a source of randomness; named after Helmut Schmidt, who pioneered the use of such machines in parapsychology, particularly in tests of psychokinesis.

SCORE. A number which represents the degree of "success" attained in a test of psi, e.g. the number of "hits" made in any given unit of trials.

SCREENED TOUCH-MATCHING (STM). A technique for testing clairvoyance using ESP cards in which, at each trial, the percipient points to one of five key positions, thereby indicating what they think the top card is in the target pack held face-down by the experimenter behind a screen. The experimenter then lays that card, still face-down, opposite the indicated position, and the check-up is made at the

end of the run.

SCRIPT. An example of automatic writing or of a record of automatic speech. (A)

SCRYING. A technique for obtaining paranormal impressions by staring into a crystal ball, pool of water, coffee grounds, tea leaves, etc., which cause images or exteriorized hallucinations. (variant of descry)

SD. See STANDARD DEVIATION.

SÉANCE. A meeting of one or more persons with a medium for the purpose of receiving communications from the deceased. Also called a "sitting" or "session". (fm. F., der. old F. seoir, "to sit", ultimately der. L. sedere, "to sit")

SECOND SIGHT. Concept used in the Celtic folklore of the supernatural, and encompassing what would today be referred to as "psychic ability".

SEGMENT. A unit of a set of trials in an experiment, often used to refer to a set of five consecutive trials in a run of ESP card-guessing.

SENDER. Less technical expression than "agent", used to denote the person or subject designated as the "transmitter" of telepathic information..

SENSITIVE. A person who frequently experiences extrasensory perception and who can sometimes induce it at will. Cf. HIGH-SCORING SUBJECT; PARAGNOST; SPECIAL SUBJECT

SERIAL-POSITION EFFECT. Expression borrowed from psychology by Michael Thalbourne ("Two long-distance ESP drawing experiments between Austria and Iceland", Research In Parapsychology, 1979, Scarecrow Press: Metuchen, N.J., 1980) to refer to a psi effect in which the level of scoring on a particular trial is related to the position which that trial holds in a series; very similar to a position effect, except that no reference is implied to the structure of the record-sheet or the subject's awareness thereof; it is assumed to be caused by psychological changes in the subject over time and/or their response to the structure of the experiment itself. See also DECLINE EFFECT; INCLINE EFFECT; TERMINAL SALIENCE; U-CURVE

SERIALITY. Universal principle, postulated by Paul Kammerer, on a level with causality, but which operates independently of physical causation and brings similar events together, as a coincidence or series. Cf. SYNCHRONICITY

SERIES. Several runs or experimental sessions that are grouped in accordance

with the stated purpose and design of the experiment. (JP)

SESSION. (i) A unit of a psi experiment, comprising all of the trials completed on a single testing occasion or sitting; an experiment may consist of one or more sessions; (ii) a sitting in a mediumistic experiment. Cf. SÉANCE (DBW)

SET. A subdivision of the record page, serving as a scoring unit for a consecutive group of trials, usually for the same target. (JP)

SHAMAN. A tribal medium, witch-doctor, priest accredited with supernatural powers, as originally exemplified by Siberian tribes. (fm. German Schamane, der. Russian shaman, der. Tungusic samân)

SHEEP. Originally, a term borrowed from a New Testament metaphor by Gertrude Schmeidler to describe a subject who does not reject the possibility of ESP occurring under the conditions of the given experimental situation; this rather narrow definition has been extended to refer also, or alternatively, to persons who believe that ESP exists as a genuine phenomenon, or even to persons who score high on various projective, scalar or checklist measures of belief in and/or experience of different sorts of putative psi phenomena. Hence also "super-sheep" (or "white sheep"), term introduced by John Beloff and David Bate ("Research report for the year 1968-69, University of Edinburgh Parapsychology Unit", JSPR, 1970, 45, 297-301) to describe a subject who is sure that their score on the ESP test will be high, by virtue of their own psychic ability. Cf. GOAT

See also SHEEP-GOAT EFFECT

SHEEP-GOAT EFFECT (SGE). Term first used by Gertrude Schmeidler to describe the relationship between acceptance of the possibility of ESP occurring under the given experimental conditions, and the level of ESP scoring, those not rejecting the possibility ("sheep") tending to score above chance and those rejecting it ("goats") at or below chance; the terms "sheep" and "goat" are nowadays often used in a more extended sense, and "SGE" may thus refer to any significant scoring difference between these two groups as defined by the experimenter.

SHIN. Term coined by R.H. Thouless and B.P. Wiesner (1947) to refer neutrally to the self, ego or soul as the agent and experient in mind/body interactions; that is, a soul-like entity involved in the processes of perception and volition;

their basic hypothesis is that Shin learns about impacts on the afferent nervous system by ESP (or "psi-gamma"), and controls the body's motor activity by PK (or "psi-kappa") on the efferent nervous system; psi is thus simply the exosomatic form of processes which are normally endosomatic: in normal perception/volition, Shin is mediated by the brain and nervous system, but in psi interactions it is directly connected to the external object or event. (Shin is the Hebrew letter ש)

SIDDHIS (also SIDHIS). Sanskrit word meaning approximately "paranormal abilities" and embracing certain creative capabilities which are generally considered quite beyond the normal repertoire of human behaviour; as described by Patanjali (circa 1500 B.C.) in his "Yoga Sutras", they include knowledge of objects hidden from view, awareness of past and future, fully developed feelings of friendliness and compassion, enhancement of sensory thresholds (e.g. "celestial hearing"), invisibility, and levitation (or "flying"); according to exponents of transcendental meditation, the modus operandi of siddhis is the experience of "sanyama", a spontaneous mental process that begins to take place once "samadhi", or "pure consciousness", is established to a sufficient degree of stability to coexist with mental activity: when this situation is present, it becomes possible for pure consciousness to "adopt" a thought, intention or desire in such a way that it is spontaneously carried out without any effort or action on the part of the individual; siddhis are seen as by-products of meditation, and not as ends in themselves but as means of developing the state of total integration known as "unity consciousness". Hence, "siddha", a person who has been instructed in the siddhis, or who has attained that level of consciousness where they begin to manifest.

SIGNIFICANCE, STATISTICAL. The improbability of the test statistic, given the assumption that the null hypothesis is true. Hence, a "test of statistical significance": the computation, by means of a mathematical formula, of a number (such as a Chi-square, critical ratio, t-ratio) which can be compared to a theoretical distribution to determine the improbability of it being due to chance alone. A numerical result is said to be "statistically significant" when the probability that it (or a result more divergent from mean chance expectation) would occur due

merely to chance factors, is equal to, or lower than, some pre-set criterion known as "alpha": alpha is traditionally (though necessarily arbitrarily) set at 0.05 (that is, odds of 19 to 1 against chance, or greater), though many parapsychologists insist on an alpha level of 0.01 and consider results which are significant at only the 0.05 level to be no more than suggestive that the null hypothesis should be rejected; if a result is determined to be "significant", then one becomes more confident that the causal relationships postulated in the experimental hypothesis are in fact at work, rather than the random vagaries of chance. A result may be termed "marginally significant" if the probability approaches the specified alpha without quite reaching it, and such a situation is often interpreted as suggesting the falsity of the null hypothesis without conclusively demonstrating it.

SII. See STUART INTEREST INVENTORY.

SINGLES TEST. A technique for testing psychokinesis in which the aim of the subject is to try to influence dice to fall with a specified face uppermost.

SITTER. A person who sits with a medium at a séance and who receives a communication through the medium. See also PROXY SITTER

SITTING. A session or interview with a medium, generally by an individual or a small number of people, and often for the purpose of obtaining communications from the deceased; also termed a "séance". See also SITTER; PROXY-SITTING

SKIN VISION. See DERMO-OPTICAL PERCEPTION; EYELESS SIGHT.

SLATE WRITING. The supposedly paranormal appearance of written messages on slates in the presence of a medium. (D&W)

SPEAKING IN TONGUES. See GLOSSOLALIA.

SPECIAL SUBJECT. Term used of a person who displays a talent for some particular paranormal ability. Cf. HIGH-SCORING SUBJECT

SPIRIT. See DISCARNATE ENTITY.

SPIRIT HYPOTHESIS. The theory that individual consciousness survives death and may be communicated with by living persons. Cf. SURVIVAL (N)

SPIRIT PHOTOGRAPHY. The purported photographing of self-portraits of discarnate entities, called "extras", upon film or photographic plates. Cf. PHOTOGRAPHY, PARANORMAL

SPIRITISM. Another and more accurate term for Spiritualism.

SPIRITUALISM. Quasi-religious cult based upon the belief that survival of death is a reality, and upon the practice of communicating with deceased persons, usually via a medium. Hence, "Spiritualist", a devotee of Spiritualism.

SPONTANEOUS. Adjective applied to any paranormal occurrence which takes place in the real-life situation -- naturally and unanticipated -- as opposed to the experimentally elicited psi phenomena of the parapsychology laboratory. Hence, for example, "spontaneous case", a discrete incident ostensibly involving spontaneous psi.

STACKING EFFECT. A spuriously high (or low) score in an ESP test which may occur when two or more subjects make guesses in relation to the same target sequence; it is due to a fortuitous relationship occurring between the guessing biases of the subjects and the peculiarities of the target sequence, and thus cannot be taken as indicating ESP; statistically, the problem is that to the extent that the guesses of the subjects resemble one another, to that extent they are awarded a similar score vis-à-vis the target, thus reducing the variance of the subjects' scores, which has an effect on the significance of the statistic in question; various statistical corrections are available to discount the stacking effect, e.g. the Greville correction and the Majority-vote technique.

STAGE TELEPATHY. Synonym for mentalism, which, as opposed to genuine telepathy, is generally accomplished by means of pre-arranged codes between "agent" and "percipient", muscle-reading, or legerdemain.

STANDARD DEVIATION (SD). A statistic which represents the degree to which a group of scores are "scattered", or "deviate from" their average; it is usually the square root of the variance; it is widely used in tests of statistical significance; in parapsychology, the term is often used, somewhat idiosyncratically, to refer to the standard deviation around the theoretical mean of a group of scores, rather than around the actual obtained mean. For the so-called "binomial distribution", the theoretical standard deviation is obtained from the formula $\sqrt{NPQ}$, where N is the number of discrete trials, P is the probability of a "hit" on each trial, and Q the probability of a miss.

STIGMATA. Term used to refer to the marks or haemorrhages which appear spontaneously on the bodies of certain devout individuals and which imitate the wounds received by Christ at the Crucifixion. (plural of Gk. stigma, "puncture, mark, spot")

STIMULUS. The object or event which causes a change in an organism; often used in parapsychology as a synonym for "target". Cf. RESPONSE

STM. See SCREENED-TOUCH MATCHING.

STUART INTEREST INVENTORY (SII). A psychological scale developed by Charles Stuart (1946) which is basically a list of 60 items, each one an activity, event or object familiar, and of possible interest, to college students; the subject rates each item on a five-point scale, ranging from "dislike very much" (-2), through "indifferent" (0), to "like very much" (+2), and their score is the sum of the 60 ratings, and supposedly measures "affectability"; Betty Humphrey ("A new scale for separating high- and low-scoring subjects in ESP tests", JP, 1950, 14, 9-23) devised a shorter version consisting of the 14 items which most clearly discriminated between "affectable" and "unaffectable" subjects as regards their ESP scores.

SUBJECT. An individual (human or animal) participating in an experiment, and of whose behaviour some aspect is being studied; in parapsychology, that "aspect" is the individual's performance on a test of psi; in a test of GESP, the term "subject" may refer either to agent or percipient, but generally means the latter, and indeed "subject" and "percipient" are often used synonymously; the term may also be applied to a person who experiences a spontaneous psi phenomenon.

SUBJECT VARIANCE. See under VARIANCE.

SUBLIMINAL. Occurring beneath the "threshold" of conscious awareness; for example, "subliminal perception": the reception, without awareness, of a stimulus; i.e., the psychological phenomenon in which a physical event (such as a word, picture, or the physiological correlate of a thought) occurring at a certain time and place,can evoke a correlated happening, be it a gesture, a dream, a spoken word or some measurable physiological change, occurring at a different time and place, and all this without any awareness by the transmitter-person or the receiver,

or indeed by any external observer, of the intermediate stages in this communication. Cf. SUPRALIMINAL (fm. L. sub, "below, under", + limen (liminis), "threshold")

SUPER-ESP HYPOTHESIS. The theory that since there are no known limits to the scope of psi, ESP in the living could theoretically be used to produce such complex phenomena as ostensible spirit communication, and that therefore the spirit hypothesis is unnecessary and unparsimonious.

SUPERNORMAL. An older term for "paranormal", the latter being preferred by parapsychologists on the grounds that it does not convey overtones of the supernatural or superiority, but merely denotes that the phenomena so referred to are different and not currently congruent with known forms of causation.

SUPER-PSI. See SUPER-ESP HYPOTHESIS.

SUPER-SHEEP. See under SHEEP.

SUPRALIMINAL. Occurring above the "threshold" of conscious awareness. Cf. SUBLIMINAL (fm. L. supra, "above", + limen (liminis), "threshold")

SURGERY, PSYCHIC. See PSYCHIC SURGERY.

SURVIVAL. Continued existence of the consciousness of the individual person in some form and for at least some time after the destruction of their physical body; life-after-death; not to be considered synonymous with "immortality", which implies unending existence. See also REBIRTH; REINCARNATION; SPIRITUALISM

SYNCHRONICITY. Term coined by Carl Jung (with W. Pauli, The interpretation of nature and the psyche, 1955) to refer to the occurrence of acausal but meaningful coincidences in time of events having the same or similar meaning to an experient. (fm. Gk. synchronos, der. synchronizein, "to be contemporary with", der. syn, "with", + chronos, "time")

TABLE TILTING. A form of motor automatism in which several persons place their finger tips on a table top, causing it to move and rap out messages by means of a code. Also called "table tipping" or "table turning". (D&W)

TACHISTOSCOPE. A laboratory apparatus used in experimental psychology, equipped with a timing device and speed shutter for controlling very brief exposures of images. (fm. Gk. tachistos, "swiftest") (W)

TARGET. In a test of ESP, the physical or mental object or event constituting the information to be paranormally acquired, or to which the subject is attempting to respond; in a test of PK, the physical system, or a prescribed outcome thereof, which the subject is attempting to influence or bring about. Hence, "target sequence", a series of targets.

TARGET CARD. In a test of ESP, the card which the percipient is attempting to identify, or otherwise ascertain.

TARGET FACE. In a test of psychokinesis, the face or side of the falling die which the subject is trying to influence so that it falls uppermost.

TARGET PACK. In a test of ESP, the pack of cards the order of which the subject is attempting to identify, or otherwise ascertain.

TARGET PERSON. The person about whom a sensitive is trying to get impressions, especially in the context of a token-object test or billet reading; the term may also be used of the subject in a test of telepathy or GESP. (D&W)

TC. See TRANSPERSONAL CONSCIOUSNESS.

TELAESTHESIA. Obsolete term coined by Frederic Myers to indicate perception, of objective events, received at a distance without the normal operation of the recognized senses; the term has been superseded by "clairvoyance". (fm. Gk. tēle, "far away", + aisthēsia, "perceptive state", der. aisthanesthai, "to perceive")

TELEKINESIS. Older term for "psychokinesis", still preferred in the USSR and Eastern Europe. (fm. Gk. tēle, + kinēsis, "a moving, disturbance") (D&W)

TELEPATHIC LEAKAGE. Expression used by Jan Ehrenwald to describe a form of telepathy which occurs between patient and therapist, such that the emotionally charged interests which preoccupy the therapist are reflected in the material produced by the patient. Cf. DOCTRINAL COMPLIANCE

TELEPATHY. Paranormal acquisition of information concerning the thoughts, feelings or activity of another conscious being; the word has superseded earlier expressions such as "thought-transference". Hence, "latent telepathy", in which there seems to be a time lag between the "transmission" or cogitation of the target and the percipient's awareness of that target. Hence also "precognitive telepathy", in which there is paranormal acquisition of information concerning the

future mental state of another conscious being. See also GAMMA-(γ)-TELEPATHY; GENERAL EXTRASENSORY PERCEPTION; ASSOCIATION THEORY OF TELEPATHY; KAPPA-(κ)-TELEPATHY; PURE TELEPATHY TECHNIQUE (fm. Gk. tēle, + pathein, "to have suffered, been affected by something")

TELEPATHY, PRECOGNITIVE. See under TELEPATHY.

TELEPORTATION. The paranormal transportation of a physical object or person, from one place to another and/or through other objects. See also APPORT (fm. Gk. tēle, + L. portatio, der. portare, "to carry")

TELERGY. Term coined by Frederic Myers (Human personality and its survival of bodily death", Longmans: Cambridge, USA, 1903) to refer to psychokinetic influence exerted over the brain and nervous system of another organism. Cf. ACTIVE-AGENT TELEPATHY; KAPPA-(κ)-TELEPATHY; MENTAL OR BEHAVIOURAL INFLUENCE OF AN AGENT (fm. Gk. tēle, + ergon, "work")

TERMINAL SALIENCE. The occurrence of higher scores at both ends of an experiment (or experimental division, such as a run or segment) than in the middle section. See also U-CURVE; SALIENCE RATIO

TEST ADMINISTRATOR. The person who actually presents the experiment to the subjects and is their main contact with the experiment. (D & W)

THETA. Term introduced by William Roll to denote whatever it is that is supposed to "survive" the death of the body, or, possibly, merely the inquiry itself into the issue of survival and phenomena related thereto. (Gk. thēta, eighth letter of the Greek alphabet, and first letter in the Greek word thanatos, "death")

THOUGHT-TRANSFERENCE. See TELEPATHY.

THOUGHTOGRAPHY. See PHOTOGRAPHY, PARANORMAL.

TM. See TRANSCENDENTAL MEDITATION.

TOKEN OBJECT. Any object associated with the target person, and held by the sensitive in order to elicit paranormal information concerning its owner; sometimes used at a séance as a proxy for an absent sitter; often also termed a "psychometric object". See also TOKEN-OBJECT READER; TOKEN-OBJECT TEST

TOKEN-OBJECT READER. A person or sensitive who gives a reading by means of a token object. Also termed a "psychometrist". See also TOKEN-OBJECT TEST

TOKEN-OBJECT TEST. A test in which the sensitive holds an object and attempts to obtain paranormal information about its history and/or events in the life of its owner; the term has in recent years come to supereede Buchanan's coinage "psychometry" (meaning literally "measurement of the soul") See also TOKEN OBJECT

TRANCE. A state of dissociation in which the individual is oblivious to their situation and surroundings, and in which various forms of automatism may be expressed; usually exhibited under hypnotic or mediumistic conditions. (fm. Old F. transe, "passage", ultimately der. L. transire, "to go across")

TRANCE PERSONALITY. See COMMUNICATOR; CONTROL.

TRANSCENDENTAL MEDITATION(TM)A meditative technique brought to the West by Maharishi Mahesh Yogi, involving the repetition of a mantra and the passive witnessing of mental activity; its effects are to produce a very deep state of physical and mental relaxation, and possibly the attainment, on the phenomenological level, of a fourth state of awareness -- "transcendental consciousness", or "samadhi" -- in which consciousness is said to achieve union ("yoga") with the Absolute ("Being"), in a state of "least excitation of consciousness", or awareness without an object, where all thinking activity is "transcended". See also MEDITATION; SIDDHIS

TRANSPERSONAL CONSCIOUSNESS (TC). A type of altered state of consciousness in which awareness of the personal self is lost in, or identified with, an awareness of the world at large or other living beings; it is said to typify certain mystical states, and to be inducible by meditation. (fm. L. trans, "across, beyond")

TRANSTEMPORAL INHIBITION. Concept introduced by Charles Tart ("Space, time and mind", in Research in parapsychology, 1977, Scarecrow Press: Metuchen, N.J., 1978) to explain the problem of how minds with (presumably) unlimited access to events past and future can cope with present reality: it is postulated that the percipient actually suppresses extrasensory information of past and future events in order to enhance ESP of the present (the evidence for this consisting of the finding of negative displacement scores on the just-previous and just-following targets); the term is an analogue of "lateral inhibition", a physiological mechanism in the

visual system which sets up boundaries that accentuate the difference between figure and ground; transtemporal inhibition likewise accentuates the difference between "then" and "now". (fm. L. trans, "across", + tempus (temporis), "time")

TRAVELLING CLAIRVOYANCE. Expression used by the 19th century Mesmerists to describe a type of clairvoyance in which the subject makes a "mental visit" to a distant place, with the percipient, upon "returning", describing what occurred there; similar to an out-of-the-body experience.

TRIAL. A unit of performance in a test of psi: in ESP tests, it is a single attempt to gain information paranormally about a target; in PK tests, it is a single attempt to influence the state of a physical system; for each trial, the subject is assigned a certain score.

t-TEST. A statistical test which yields a statistic (called a "t-ratio") similar to the critical ratio (or z-score) but which, unlike the latter, is based upon the actual number of scores making up the sample of data: (i) a single-sample t-test is used where there is just one group of scores under consideration, and a test is required to determine whether the observed average deviates significantly from mean chance expectation; (ii) an independent-samples t-test is used to determine whether, for two different groups of subjects (e.g. males versus females), the two mean scores may be considered as being significantly different from one another, or as being essentially identical (that is, as being "drawn from the same population"); and (iii) a correlated (or dependent) samples t-test is similar to ii), except that the measurements consist of two sets of data taken from the same group of subjects (e.g. in a before/after design), or from pairs of subjects that have been matched in some way.

TWO-TAILED TEST. A test of statistical significance is said to be "two-tailed" if the direction of the descriptive statistic has not been specified as being expected to be positive (i.e. in the above-chance region), or negative (i.e. in the below-chance region). Cf. ONE-TAILED TEST

TYPE I (ALPHA) ERROR. Term used to describe a kind of error made in the course of statistical inference, viz. the rejection of the null hypothesis when it is in fact true (i.e. deciding that the results are due to a specific cause whereas in

fact they are due to nothing more than chance). Cf. TYPE II (BETA) ERROR

TYPE II (BETA) ERROR. Term used to describe a kind of error made in the course of statistical inference, viz. acceptance of the null hypothesis when it is in fact false (i.e. deciding that the results are due to nothing apart from chance whereas in fact they are due to some specific causal factor.) Cf. TYPE I (ALPHA) ERROR

U-CURVE. A scoring trend that displays initially high scores, followed by a decline, terminating in a recovery to high scores, such that a graph of the scoring resembles the shape of a "U"; hence also an "inverted U-curve", where initially low scores are followed by a rise to high scores, but terminating in a return to low scores; in mathematical terminology, each U-curve may be represented by a "quadratic function". See also TERMINAL SALIENCE; SALIENCE RATIO. Cf. DECLINE EFFECT; INCLINE EFFECT; SERIAL-POSITION EFFECT.

UNORTHODOX HEALING. See HEALING, PSYCHIC.

UP THROUGH (UT). A technique for testing clairvoyance in which the subject guesses the order, bottom to top, of a pack of ESP cards before any are removed or compared with the calls. Cf. DOWN THROUGH

UT. See UP THROUGH.

VARIANCE. A statistic which represents the degree to which a group of scores are "scattered" or "dispersed" around their average; formally, it is the average of the squared deviations from the mean; in parapsychology, the term is often used somewhat idiosyncratically to refer to the variance around the theoretical mean of a group of scores, rather than around the actual, obtained mean. Hence, "high variance": the tendency of scores to fluctuate around the mean to a degree greater than that expected by chance; "low variance": the tendency of scores to fluctuate around the mean to a degree less than that expected by chance; "mean variance (theoretical)": the expected variance of the theoretical mean score; "run-score variance": the variance, around the theoretical mean, of the scores obtained on individual runs; "subject variance": the variance, around the theoretical mean, of a subject's total score. (JP)

VARIANCE-DIFFERENTIAL EFFECT. A statistically significant difference between the variances of the scores of runs (or of other units) in two experimental sessions designed to affect results differentially. (J.P.)

VERIDICAL. Truthful; corresponding to, or conveying, fact; used with reference to items obtained paranormally which have been verified. See also DREAM, VERIDICAL HALLUCINATION (fm. L. veridicus, der. verum, "truth", + dicere, "to say") (A)

VERIFICATION. See AUTHENTICATION.

VOICE PHENOMENA. See ELECTRONIC VOICE PHENOMENA.

WAITING TECHNIQUE, Term used by John Beloff and Ian Mandleberg ("An attempted validation of the "Waiting Technique", JSPR, 1967, 44, 82-88) to describe a method of responding to ESP targets in which the percipient deliberately waits for images to develop , as described by Rhea White ("A comparison of old and new methods of response to targets in ESP experiments", JASPR, 1964, 58, 21-56).

WATER DIVINING. See DOWSING.

WATER WITCHING. See DOWSING.

WHITE SHEEP. See under SHEEP.

WITNESS EFFECT. A change in scoring level during a test of psi which sometimes occurs when a stranger is brought into an experimental session in the role of observer.

WORD ASSOCIATION TEST. A psychological test, originally devised by Carl Jung, in which a person is required to respond to a list of stimulus words with the first word that comes to mind; these"associates", and the length of time taken before they are produced, are said to indicate aspect's of the individual's personality structure; used by Whately Carington to investigate the personality of mediums and their controls.

XENOGLOSSY. Term coined by Charles Richet ("Xénoglossie: L'écriture automatique en langues étrangères", PSPR, 1905, 19, 162-194) to denote the act of speaking in a recognized foreign language ostensibly unknown to the speaker. Hence, "recitative xenoglossy": the mere utterance of fragments of the language, as from rote memory;

"responsive xenoglossy": the ability to converse more or less freely in the unknown language. To be distinguished from glossolalia. (fm. Gk. xenos, "foreign, alien", + glōssa, "language")

ZENER CARDS. The original name given by J.B. Rhine to the ESP cards; named after the perceptual psychologist Karl Zener, a colleague of Rhine's, who apparently suggested the symbols to be used on the cards.

z-SCORE. See CRITICAL RATIO.